

Albright's last scientific stand

The US Secretary of State has announced some tentative steps towards the elusive goal of a scientifically literate State Department. Her successor will need to go further.

Scientific leaders in the United States have been whistling in the dark for many years now about the diminishing scientific and technical capabilities of the State Department, and were obviously relieved to hear Madeleine Albright, the Secretary of State, acknowledge their complaint directly in an address to the annual meeting of the American Association for the Advancement of Science (AAAS) last week.

The drumbeat of complaint about the scientific competence of the State Department has been driven by a rising number of science- and technology-related issues, ranging from trade in genetically modified foods to the suspected proliferation of biological weapons, which confront today's diplomatic corps. In the United States this has coincided with a diminishing of that corps. In a report requested by the department, and delivered in preliminary form in 1998 and in its final form last year, the National Research Council (NRC) called for extensive remedial action.

"The State Department's scientific capabilities have not always been as substantial as they should be," said Albright to the AAAS. Her speech promised a policy statement on science, to appear this month, the appointment of a senior science adviser to the secretary (something mandated last autumn by the Congress) and the appointment of a science counsellor in New Delhi.

Beyond that, Albright had little concrete to offer. "This is not the time and place for an extended response to the NRC's report," the secretary said, leaving some of the audience wondering what would be a better time or place for such a response than a AAAS meeting in Washington, more than a year after the bulk of the NRC's advice was delivered. The speech made no mention of the scientific advisory panel

recommended by the NRC, or of its sensible suggestion that the State Department tap scientific expertise from other government agencies. It also hinted that no extra resources would be available from the State Department's existing budget to bolster its scientific expertise.

Despite the dearth of specifics in Albright's address, officials who have met with her say that she is genuinely committed to the rebirth of science at the State Department. But Albright's tenure is liable to end next January, with President Clinton's. That is one reason why scientific leaders have been working closely with the State Department's permanent bureaucracy to build support for reform. Optimists among them say that important civil servants now support the necessary change. They will need to do so strongly if the current momentum for reform is to survive the change of administration.

A task force inside the department is now drawing up a science-policy document and, more importantly, an action plan to strengthen its scientific capabilities. The science adviser to the secretary will probably be appointed this summer, to serve out the closing months of the Clinton administration.

It will then fall to Albright's successor to implement the action plan. There is no guarantee that the vital but somewhat obscure issue of scientific competence will capture that person's attention. It took Albright 18 months to find time in her diary — early last month — for a meeting requested by Bruce Alberts, the president of the National Academy of Sciences; and Albright's engagement of this issue has occurred three years into her term of office. If her successor takes as long, last week's speech will be remembered less for its worthy sentiments than for its failure to announce meaningful action to update the State Department's approach to science and technology. ■

A critical nuclear appointment

The resignation of a chief executive poses a challenge for a profitable international company in desperate circumstances.

"Where science never sleeps" — so runs the catch-phrase of the public relations campaign for British Nuclear Fuels Limited (BNFL). And the science is extensive and challenging — instrumentation in hostile environments, technical evaluation of radioactive fuel inventories, Monte Carlo simulations for criticality safety margins and shielding, modelling of fuel elements in all possible conditions, magnetic technologies for separating particles in gases and levitated bearings, magnetohydrodynamics of glass vitrification, and much more. But some of the outcomes have been problematic: after years of delay before getting approval in 1994 to begin operations, the Thorp reprocessing centre has been severely hit by erosion and blockages in feeder lines.

Whatever the quality of the engineering, the management and culture at BNFL have been catastrophic for its reputation and for the profile of the nuclear industry as a whole. Important overseas customers in Japan and Germany have been outraged to discover falsified safety documentation. Investigating those scandals, the UK Nuclear Installations Inspectorate earlier this month condemned the company for its systematic management failures and lack of a safety culture.

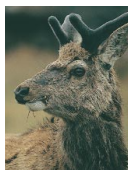
BNFL is wholly owned by the UK government, which would dearly love to privatize half of it — with £2 billion of annual turnover and profits running at about 10 per cent, it could be an attractive proposition under the right leadership. And it has a future: last year it bought Westinghouse, clinching its leading role in decommissioning and waste management in the United States and elsewhere, as well as supplying those countries in Asia still willing to invest in new nuclear plant.

In a recent study the Royal Society concluded that the nuclear option needed to be retained as other sources of energy run out or come under increasing question. But even if nuclear power is phased out altogether, its 'back end' will be with us for a long time. For those reasons, not only the board of BNFL but also the British government have a critical responsibility in the next few days or weeks in appointing a successor to replace John Taylor, the BNFL chief executive who resigned this week. That person will need grasp, acute sensitivity to the public interest, and a persona, to transform the culture and image of a business that will need constantly, in many countries, to demonstrate its absolute integrity, in both technical and moral senses. ■

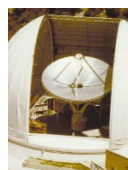




Fossil dealing
Australian jailed for
taking dinosaur and
human footprints
p4



Red deer
Culling plan could
jeopardize 30 years
of research
p5



Radio astronomy
12-metre Kitt Peak
telescope faces the
final curtain
p7



Origin of AIDS
Are polio vaccines
made from monkeys
to blame?
p9

Germany challenges human stem cell patent awarded 'by mistake'

Munich

Germany's federal minister of legal affairs, Herta Däubler-Gmelin, announced last week that she is to file a formal objection against a patent issued by the European Patent Office (EPO) that includes claims on technologies that could be used to alter the composition of the human germ line.

The move follows an admission by patent office officials — prompted by an investigation by the environmental group Greenpeace and the *Financial Times Germany* — that “a very serious error” had been committed by the EPO when it granted the patent in December last year.

The patent violates the EPO's own rules as determined in the European Patent Convention, according to which patents must not violate common standards of ethics and morality. It also conflicts with a European Union (EU) patenting directive, although this is not legally binding in this case.

The EPO pointed out in a statement that it does not have the legal right to amend a granted patent containing an error “on its own initiative”. But it added that the European Patent Convention allows anyone to challenge a patent by filing an opposition within nine months after it is granted.

The patent, covering the “isolation, selection and propagation of animal transgenic stem cells” (EP 0695351 B1), was filed in 1994 by the University of Edinburgh's Centre for Genome Research, and the techniques were subsequently licensed exclusively to the Australian company Stem Cell Sciences (SCS).

The patent covers techniques for generating transgenic animals from stem cells. In principle, it relates to methods for purifying specific types of genetically engineered stem cells from mixtures of cultured cells.

The patent has provoked controversy by not explicitly excluding the application of such techniques to humans; indeed, it states that the term ‘animal cell’ “is intended to embrace all animal cells, especially of mammalian species, including human cells”.

But SCS denies “any interest or intention to patent or develop technologies for human genetic engineering”. According to Peter Mountford, the company's chief scientific



Life guards? Greenpeace last week bricked up the entrance to the EPO in protest at the patent.

officer, SCS's main interest in the technology is in developing “a wide range of cell-based assays for [the] pharmaceutical industry”.

“Our techniques may have been misunderstood as being intended for use in methods for producing genetically engineered

humans,” Mountford said in a statement that acknowledged that “the interpretation of one granted claim could extend to a method for genetically altering humans”.

The potential applications of genetic engineering to changing the germ line are a particularly sensitive topic in Germany, partly for historical reasons. In a parliamentary debate last week, all of the country's political parties dissociated themselves from the patent.

Many participants expressed concern that the issuing of the patent showed that Germany's strict bioethical standards could be undermined by commercial pressures. Scientists, however, are worried that the EPO's mis-handling of the patent could further inflame public distrust in genetic research and in the activities of Germany's relatively young biotechnology industry.

Germany's research minister, Edelgard Bulmahn, criticizing the lack of transparency in the EPO's patenting procedures, called on the convention's 18 signatory states to “fundamentally reconsider” the patent office's

UK university recruitment in danger

London

British vice-chancellors warned last week that academic recruitment difficulties — especially in science — could undermine the future of the United Kingdom's ‘knowledge economy’.

A report commissioned by the Committee of Vice-Chancellors and Principals says that academia is finding it increasingly difficult to recruit high-calibre professors and senior academics. Based on a survey of 13 universities, the report says that recruitment problems have caused delays in teaching and research.

The report also highlights the fact that posts are being held vacant, and part-time and visiting lecturers increasingly used to cover for long-term vacancies. Scientific disciplines, such as engineering, computer science, maths, biology and chemistry, feature in the list of problem areas.

Meanwhile, a separate report by the Association of University Teachers (AUT) — the main labour union for UK university staff — warns that the advancing age profile of university academics signals a looming recruitment crisis.

The lack of high-calibre candidates to replace them is said to be due to poor pay and career prospects at the early career stage. One former academic, now a pharmaceutical research scientist, says her move was ultimately prompted by the long hours and increasing administrative and technical responsibilities. But she was also upset that her university salary was lower than that of her postdoctoral worker.

David Triesman, general secretary of the AUT, is angry at the long-term neglect of recruitment issues. “In the sciences I can't see on present dynamics how people will be replaced.”

Natasha Loder

rules, and in particular its appeals system. Unlike the US Patent and Trademark Office, where appeals are heard in a federal court, the European appeals system is based on in-house 'technical' hearings.

The patent conflicts with a new EU directive on patenting biotechnological processes, which excludes genetic modifications of human embryonic stem cells from being patented. But although the EPO belongs to an organization of 19 European states, it is not part of the EU, allowing the patent — in principle — to be legally valid. The EPO has agreed, however, to adapt the directive to its own rule of behaviour.

The EPO immediately conceded its error after the case became public last week. According to Christian Gugerell, director of biotechnology at the EPO, the application had not been investigated thoroughly enough. "The English wording of the claim should have included the qualification 'non-human', because in English scientific usage the term 'animal' [is normally taken to] include 'human'," he says.

But Christoph Then, principal campaigner against genetic engineering at Greenpeace Germany, is sceptical about the EPO's claim that the mistake was merely the result of a lack of thought. "Given the explicit wording of the patent claim, it is out of the question that the application has simply 'slipped through,'" he says.

Greenpeace activists last week bricked up the entrance to the EPO, blocking access for several hours. They hope the case will boost worldwide protests against the patenting of life forms. Controversial applications currently being processed include the technology used to produce 'Dolly' the sheep (see *Nature* 403, 351; 2000).

Quirin Schiermeier

US energy agency pulls plug on role in genome project

Washington

Scientists at the US Department of Energy (DoE) say they will stop large-scale human gene sequencing later this year, when a preliminary draft of the human genome sequence is due to be completed.

The department took the lead in initiating the Human Genome Project in the mid-1980s, but has since become a minority participant in what is now a global project.

The Joint Genome Institute (JGI) — a powerful sequencing centre established by several DoE laboratories at Walnut Creek in California — will continue with limited sequencing to refine the human genome data, but is planning a new work programme for 2001.

"We see ourselves as being out of the human genome and the bulk of the mouse genome by October or so," says Elbert Branscomb, JGI's director.

The DoE's division of biological and environmental research, which runs its component of the genome project, will meet with its genome contractors this week in Santa Fe, New Mexico, to begin planning the division's new programme.

Ari Patrinos, head of the division, says that the new research effort will focus on proteomics and structural biology, work which will rely on two areas where DoE laboratories are stronger than most universities: large facilities — such as high-powered light sources — and computational science.

Some DoE scientists claim the move is

driven in part by the department's exasperation at the extent to which its main US partner, the National Human Genetics Research Institute (NHGRI) at the National Institutes of Health (NIH), has claimed most of the credit for the project.

NHGRI has received large budget increases, driving the DoE share of the project — worth about \$90 million a year — down from an initial one-third of the US project to about one-fifth. But Patrinos says the DoE will maintain its investment at \$90 million "for the next couple of years" and collaboration with NHGRI will continue.

Patrinos acknowledges that some individuals at the DoE are worried that the final summary paper publishing the draft sequence will fail to acknowledge the department's role in the project. But he says that he is "philosophical" about this prospect.

As the relationship with NHGRI weakens, Patrinos hopes to strengthen links with two other NIH institutes, the National Cancer Institute and the National Institute of General Medical Sciences.

The JGI hopes to move on to study other genomes related to human health and environmental problems.

Branscomb says that up to a quarter of its capacity may be used to sequence microbes that have the capacity for bioremediation or for sequestering atmospheric carbon. He also hopes to win support for the sequencing of other vertebrate genomes for comparison with the human sequence.

Colin Macilwain

Australian jailed for removal of fossilized footprints

Sydney

Australian authorities have sent a strong warning to those who trade illegally in fossils by jailing a man for two years for removing and trying to sell a dinosaur footprint and two human footprints.

Michael Latham, an Aborigine from the town of Broome, Western Australia, confessed that he had used an angle grinder to cut a block containing the dinosaur footprint when it was exposed at low tide. On a disability pension and needing money to support his family, he offered the footprint to a collector for A\$250,000 (US\$153,000), but later dropped the price to A\$100,000.

He also extracted two fossilized human footprints from another site, thought to be 7,000 years old. A local cattle farmer, accused of receiving these, was acquitted.

John Long, an expert in dinosaur fossils and Curator of Vertebrate Palaeontology at the Western Australian Museum, Perth, says the *Megalosauropus* footprint is about 120



Footprints instead of fingerprints: police display the retrieved *Megalosauropus* fossil.

million years old, and was probably made by a carnivorous dinosaur up to nine metres long and similar to *Allosaurus*. "Broome is one of the world's most significant sites for dinosaur footprints of the early Cretaceous, but there is no protection for them," says Long (see *Nature* 403, 687, 689–690; 2000).

Such fossils have great spiritual significance to local Aborigines — who expressed their displeasure by throwing one of the human footprints back into the sea. Sentencing Latham, Judge John Wisbey said: "I'm in no doubt you were well aware of their importance to your people".

The state of Western Australia is considered the legal owner of the dinosaur footprint, since it came from below the high-water mark. It is being kept in the police station until its future location has been agreed.

Peter Pockley

NIH tightens up monitoring of gene-therapy mishaps

Washington

The US National Institutes of Health (NIH) is telephoning every investigator who has registered a clinical trial in gene therapy, in an expanded effort to track down unreported adverse events. Its results will be made available on the Internet by the end of the year, in a bid to respond to public concerns.

The move represents a shift of policy for the NIH. The Recombinant DNA Advisory Group first proposed a database of adverse effects of gene therapy in 1996. But this failed to materialize after Harold Varmus, then NIH director, removed much of the group's regulatory authority for gene therapy.

The telephone campaign reflects concern over the extent to which investigators have been ignoring the rules on reporting. Gene therapy researchers must report all adverse events — ranging from fever to death — to both the US Food and Drug Administration (FDA) and the NIH.

But the requirement has frequently been ignored. While most have reported adverse events to the FDA, which keeps such information confidential, few have submitted them to the NIH, which makes it public.

The issue was highlighted by the death of 18-year-old Jesse Gelsinger at the University of Pennsylvania last September (see *Nature* **401**, 517; 1999). The NIH subsequently sent requests for information to all institutions conducting research in adenoviral vectors — the type of vector received by Gelsinger, and used in about 25 per cent of active gene therapy trials.

A month later, the agency wrote to all institutions that have participated in any gene therapy clinical trials, regardless of the vector used. Hundreds of unreported events came to light, including several deaths not directly attributable to the experimental treatment.

Lana Skirboll, NIH associate director of science policy, says that the telephone calls are revealing still more adverse events, almost six months after the NIH sent its first letters asking for them. She adds that she is surprised by the level of non-compliance. Many investigators have said they thought they only had to notify the FDA.

Officials from both NIH and FDA were grilled at a Senate hearing last month where William Frist (Republican, Tennessee) said that the two agencies had not communicated enough about gene therapy's adverse events in the past (see *Nature* **403**, 583; 2000).

The new commitment to reporting and disseminating information about adverse effects appears to be a response to congressional and public pressure, after high-profile

stories appeared in the popular press. But some researchers question whether the reaction is appropriate, since the vast majority of these deaths are probably due to the patients' underlying disease.

Dusty Miller, for example, of the Fred Hutchinson Cancer Research Center in Seattle, applauds the information gathering. But, he asks, "How many of these adverse events are truly related to gene therapy?"

Inder Verma, professor of genetics at the Salk Institute in La Jolla, California, says that the public's response is a predictable result of releasing information that was once kept private. But he acknowledges that the abundance of information — as well as the heightened scrutiny caused by Gelsinger's death — has its disadvantages.

Newspaper reports that 20 patients in a cancer gene therapy trial were exposed to HIV, for example, turned out to be inaccurate. "The more reporting that is done, the better," says Verma, president-elect of the American Society of Gene Therapy. But he remains concerned about the form in which the NIH makes the data available.

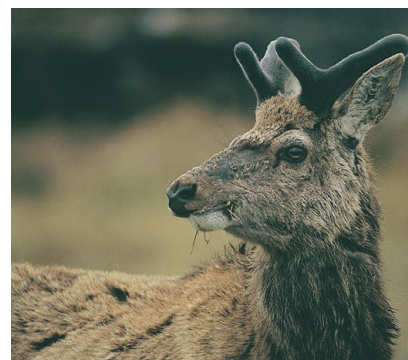
"Someone has to put the data together in a meaningful manner," says Verma. Many people feel that a properly designed database would allow researchers and patients to compare the effects of different vectors and doses in the same organ, helping them to weigh up the risks against the benefits.

Skirboll pledges that the NIH will have such a database up by the end of the year. But it remains unclear whether it will contain only adverse events directly related to gene therapy, or all adverse events that take place during or following the trial.

The field is in disagreement over this point, says Miller. Pharmaceutical and biotech representatives have repeatedly pushed for databases that only contain the adverse events directly attributable to gene therapy. Miller prefers a database containing all adverse events, but acknowledges that would be "an overwhelming amount of data".

Without proper organization, researchers could be deluged with information that they may be unable to turn into knowledge, says Verma. He adds that a properly organized, clearly accessible database could also help win back the public trust needed to get new clinical trials launched.

Paul Smaglik



Rum deal: deer face death and research faces disruption in the name of conservation.

Culling plans put future of red deer study in jeopardy

London

A world-renowned study of red-deer population dynamics on a Scottish island is being threatened by a culling proposal.

Red deer on Rum, off Scotland's west coast, have been studied by biologists for three decades in what is considered to be one of the world's most significant long-term studies of mammal populations.

But the island's owner, Scottish Natural Heritage (SNH), which is responsible for safeguarding and enhancing the "natural, genetic and scenic diversity" of Scotland, is considering reducing deer numbers from 1,500 to around 300–400.

"Our long-term aim for managing Rum is restoration and conservation," says Michael Usher, SNH's chief scientist. "That does not sit happily with a very large deer population."

At present, SNH culls only 200 head of deer a year on the island outside the study area. It now wants to restore the natural balance of wildlife, open-ground vegetation, native woodland and scrub. To help achieve this, woodland lost through overgrazing would be recreated.

Additional pressure results from the fact that Rum has been proposed as a special conservation area, under a directive on natural habitats issued by the European Commission. Under this directive, SNH is obliged to bring the features of interest into favourable conservation condition. "At the moment we would argue they are not, because they are overgrazed," says Usher.

The deer that most interest the population biologists live within a small, currently uncultured region of the island known as the 'north block'. This part is not under immediate threat — new plans for culling and tree planting will begin in



Verma: database needs right design.

▶ the south, with the northern population remaining safe for at least another five years.

Scientists are concerned that the plans could herald the end of an internationally valuable long-term population study, and hope that SNH will consider leaving the northern-block population alone.

"The long-term red-deer study has profoundly influenced our understanding of important conceptual issues in evolutionary biology — for example, sexual selection and sex allocation — and population biology, including factors such as density dependence and the maintenance of genetic variation," says Andrew Cockburn, a behavioural ecologist at the Australian National University in Canberra, who is familiar with the studies.

"The research also has had great consequences for our understanding of the management of ungulate populations. Perhaps most important, the study has clearly not run its course, as fascinating recent results on sex ratios attest. It would be very disappointing to see the work come to an end."

Josephine Pemberton, a population biologist at the University of Edinburgh is one of the authors of a recent paper on sex ratios in Rum red deer (see *Nature* 399, 459; 1999). She says the deer, which are individually monitored, "should be seen as a national and international resource for research and education which has its own intrinsic value".

Usher acknowledges the researchers' concerns. But he argues that they should now be addressing the substantial opportunities for researching how a deer population at normal density would change in a more diverse landscape.

Another scientist suggests that Rum could be treated as a demonstration case for other areas of Scotland where similar replanting is occurring next to high-density deer populations. "You would have to accept high-density populations of deer in areas adjacent to where conservationists want to plant," says Steve Albon, from the Institute of Terrestrial Ecology in Banchory, Scotland.

Although SNH says it is committed to the concept of Rum as an outdoor laboratory, it wants to limit future research. A spokesperson said "we do have a difficulty with current research work, which is conflicting with the management of the reserve". SNH's vision includes increasing visitor numbers to the island and, in the long term, allowing more people to live there.

Natasha Loder

Young, worldly and unhelpful all miss out on data sharing

Washington

Researchers who refuse to share data may provoke others to withhold results from them, according to a study by health-policy analysts at Harvard Medical School.

The study found that young researchers, those who publish a lot, and investigators seeking patents are most likely to be denied access to biomedical data and reagents. It also found that researchers who withhold data gain a reputation for this and have more difficulty in obtaining data from others.

The study, whose results are published in the February issue of *Research Policy*, was conducted by a research team led by sociologist Eric Campbell at Harvard Medical School's Institute of Health Policy.

The team surveyed 2,366 randomly selected scientists — both clinical and non-clinical — at 117 US medical schools. Overall, 12.5 per cent said they had been denied access to other academic investigators' data, excluding article reprints, during the past three years. This corresponds with previous findings by the team and other groups.

But, by examining the 'victims' of data withholding, the team identified those experiencing the most difficulty. For junior staff members, the team found that 13.5 per cent were denied access, compared with 5.1 per cent of senior researchers.

The relationship between data withholding and researchers' publishing records during the preceding three years was striking: 7.7 per cent of those who had published 1–5 articles had had data withheld from them, but this rose to 28.9 per cent for researchers who had published more than 20.

Among those who had applied for a patent, 30 per cent had been refused access to data, compared with 9.4 per cent of those

Researchers denied access to data over the past three years

Researcher	Percentage of requests denied
Has MD	9.2
No MD	17.7
Publishing record:	
1–5 articles	7.7
>20 articles	28.9
Applied for patent?	
Yes	30.0
No	9.4
Member of federal review board/study panel?	
Yes	20.2
No	8.6

who had not applied for a patent. Scientists who reported refusing to share data were more than twice as likely to be the victims of data withholding as those who had not.

"Selectively withholding research results from the most productive and commercially active researchers could slow the progress towards understanding the causes and cures of human disease," warns Campbell.

The National Institutes of Health issued guidelines last year encouraging the sharing of materials and data, but some scientists say these need tightening. Campbell and his colleagues make several recommendations to institutions regarding data sharing. For example, they suggest that senior staff should be encouraged to help junior scientists gain access to results and materials from other researchers and that professional societies and funding agencies should encourage data sharing.

"We welcome the study," said Jonathan Knight, associate secretary of the American Association of University Professors. "The study shows some concrete dimensions to the problem."

Rex Dalton

US eases Israelis' lab access

Jerusalem

Israeli scientists will have an easier time visiting and working with laboratories run by the US Department of Energy under an agreement signed last week.

Under the terms of the agreement, signed by US Secretary of Energy Bill Richardson and Israel's Minister of National Infrastructure, Eli Suissa, Israeli scientists will no longer have to undergo thorough and lengthy security clearances before being allowed to visit the laboratories.

These clearances were imposed on Israeli scientists in the wake of the conviction of

Jonathan Pollard, an American who spied for Israel, in 1986. They were tightened up in 1998 following the arrest of Wen Ho Lee, a Los Alamos scientist who is suspected of spying for China.

At their meeting last week, Richardson and Israeli officials also discussed cooperation on technical means of monitoring the Nuclear Non-proliferation Treaty.

Israel has declined to sign the treaty and has maintained a deliberate silence about its nuclear weapons programme. However, the country is widely believed to have built up a nuclear arsenal.

Haim Watzman

Promega and Roche take up battle over PCR patents

San Francisco

The Swiss-based pharmaceutical company Hoffman-LaRoche is facing a series of legal hearings over the validity of its two key patents on the polymerase chain reaction (PCR). The process will culminate in a trial in a federal court in October next year.

Last week, a judge in San Francisco announced the timescale for the proceedings, following a ruling that Roche's patent on the enzyme *Taq* polymerase — an essential component of PCR described in the patent application — had been fraudulently obtained (see *Nature* 402, 709; 1999).

The proceedings are the latest stage of a long-running legal battle between Roche and the Wisconsin-based biotechnology company Promega over issues associated with PCR, which is used to amplify small quantities of DNA. Next year's trial will also address Roche's allegation that Promega, by selling its own unlicensed native *Taq*, has induced researchers to violate Roche patents and licence agreements.

According to Brenda Furlough,

Promega's general counsel, the company's challenge to Roche's PCR patent is based on its claims that the US Patent and Trademark Office (PTO) was not told about certain previous discoveries when the PCR patent was sought more than a decade ago.

But Melinda Griffin, general counsel for Roche Molecular Systems in Alameda, California, says that Promega's claims are groundless, and that the firm expects to defeat all of Promega's challenges, and win its patent-infringement claims against the Wisconsin company.

Court records show that Promega will make its initial disclosures in April this year on what 'prior art' it claims is involved in the disclosures to the PTO that it is challenging. Furlough declines to provide details of the evidence that it intends to produce.

The PCR patents in dispute were acquired by Roche when the company purchased the Cetus Corp. of California in 1991, where Kary Mullis was working when he invented PCR — for which he won a

Nobel prize. In a previous unrelated legal battle, Roche fought the DuPont Corp. over the rights to PCR.

Promega's challenge, and Roche's infringement claims against Promega, were held in abeyance while the two companies battled over the *Taq* patent. Last December, US District Judge Vaughn Walker ruled that a Roche patent for native *Taq* was obtained through eight instances of fraud.

If Walker's ruling that the patent is 'unenforceable' stands up on appeal, Roche's patent for native *Taq* would be invalidated. At a hearing last week, Walker ruled that Roche can go forward with an expedited appeal on the ruling, despite Promega's call for other issues to be decided first. "We are very pleased the judge granted our request for an appeal now," says Griffin.

The appeal will be addressed to the Court of Appeals in Washington DC. If the court agrees to hear the appeal, a decision could take 18 months.

Roche's Thomas White said that he was confident the defence of the recombinant *Taq* patent would succeed.

Rex Dalton

Funding woes spell doom for US radio dish

Washington

One of radioastronomy's venerable telescopes — and the only millimetre-wavelength instrument that serves the entire US research community — will be shut down in July, the victim of continuing tight budgets for the National Radio Astronomy Observatory (NRAO).

The 12-Meter Telescope on Kitt Peak, Arizona, was built in 1967 and upgraded in 1984. It was a pioneering instrument in the study of interstellar molecules, and is still used by more than 150 visiting scientists a year, with staff on hand to assist with observations.

The NRAO, which runs several US radioastronomy facilities for the National Science Foundation, had hoped to keep it operational until the far more powerful Atacama Large Millimeter Array (ALMA) comes online later this decade. But years of flat funding have strained resources to the point where keeping the older telescope going is no longer an option, says Paul Vanden Bout, director of the NRAO, which is based in Charlottesville, Virginia.

Although Vanden Bout says the decision to close the telescope "wasn't strictly a budgetary matter", he admits that "the scope of our activities has exceeded our resources". A total of 24 NRAO staff in Tucson will be affected by the closure. About half are



Turn off, tune out: Kitt Peak's 12-Meter Telescope will close before its replacement is ready to open.

expected to shift over to work on ALMA, but others will be out of a job, he says. The move is expected to save around \$2 million a year in operational costs.

The NRAO, which also operates the Very Large Array in New Mexico and the Very Large Baseline Array of ten dishes located in different states, has requested \$32.5 million

for next year, the same as last year. With ALMA on the way, a flat budget would be difficult enough to manage. But the NRAO has another cloud hanging over it in the form of a \$29 million claim by COMSAT, the contractor building the new Green Bank Telescope in West Virginia (see *Nature* 384, 505; 1996).

COMSAT has charged that cost overruns on the telescope are the fault of the NRAO and its parent organization, Associated Universities, Inc. The case has gone to binding arbitration, with a decision expected this autumn. If the NRAO is forced to pay a substantial amount of the \$29 million, it would be a severe financial blow.

Vanden Bout says he is confident that NRAO lawyers made a compelling case to the arbitrators, and denies that the decision to close the telescope had anything to do with bracing for an unfavourable decision in the COMSAT dispute. The long-delayed \$74.5 million Green Bank Telescope is finally nearing completion, he says, with operation scheduled to begin by late summer or autumn.

Radioastronomers, meanwhile, are mourning the loss of the 12-metre telescope. Carl Heiles of the University of California at Berkeley calls it a "real blow", and worries that it will make it even more difficult to attract and train graduate students in millimetre-wavelength astronomy. **Tony Reichhardt**

NRAO

Canadian budget boost for science

Montreal

Paul Martin, Canada's Finance Minister, this week unveiled a budget that will cut taxes for both individuals and corporations by Can\$58 billion (US\$39.6 billion) over the next five years. The move — intended to stem the 'brain drain' to the United States — will be accompanied by a rise in spending on research and innovation to more than Can\$4.1 billion next year.

Two scientific organizations are created in the budget. Genome Canada has Can\$160 million to fund five genome science centres. These will provide laboratory services and technology to researchers from government, universities and the private sector.

The Canadian Foundation for Climate and Atmospheric Sciences will provide Can\$60 million in block grants over six years to climate-science institutes and universities, as well as peer-reviewed research grants.

In addition, Can\$900 million over five years goes to the Canada Foundation for Innovation, which provides infrastructure grants, bringing the government's total investment to Can\$1.9 billion. A Can\$900 million contribution over five years will establish 2,000 university research chairs (see *Nature* 401, 731; 1999).

Other initiatives include Can\$210 million over three years for the Climate Change Action Fund, established in 1998 to help Canada meet its international commitments, and for other energy-efficiency and renewable-energy schemes.

The Canadian International Development Agency will get Can\$100 million over four years to help developing countries reduce greenhouse gases and promote sustainable development. And Can\$90 million will be spent over three years to create a national strategy on endangered species, including legislation (see *Nature* 401, 418; 1999).

Martin's tax-reduction plans are funded from a Can\$6 billion surplus. But some argue that the extra science and technology funding lacks a clear focus. "There's no big strategy," says John de la Motte, a science-policy analyst at the University of Ottawa. "It's putting money back to where it was taken from" by recent spending cuts. **David Spurgeon**

Robert May to head Royal Society

London

Sir Robert May, currently chief scientific adviser to the British government and head of the Office of Science and Technology, is to be the new president of the Royal Society. He was elected this week to succeed the current president, Sir Aaron Klug, when the latter's term of office expires at the end of November.

May, a physicist by training who has

since become a world-renowned mathematical biologist, is professor of zoology at the University of Oxford, and was appointed to his government post in 1995 (see *Nature* 375, 527; 1995).

He has played a key role in reforming the government's science advisory procedures in the wake of the bovine spongiform encephalopathy crisis. May's five-year contract expires at the end of August.

Post-Cold War needs 'new forms of scientific linkage'

Paris

New international links between scientists are needed to meet the political realities of the post-Cold War world, a meeting in Paris was told last weekend.

Speakers at the meeting, held on the changing political role of scientific collaboration, emphasized the need to create and strengthen scientific networks in areas such as the former Yugoslavia or North Africa, where political tensions are high. They suggested that countries with experience of creating such networks should help.

Participants included seasoned experts in science and public policy, among them Alexander King, former director-general for scientific affairs for the Organization for Economic Cooperation and Development (OECD), and William Nierenberg, former deputy secretary-general for science at the North Atlantic Treaty Organization (NATO) and subsequently director of the Scripps Institution of Oceanography in California.

The conference was organized by Futuribles International, a non-profit think-tank based in Paris, and the George C. Marshall Institute. Scientists from North Africa, Eastern Europe and the Middle East were also present.

Several speakers described the influential role of scientists on the international scene during the period between the end of the Second World War and the start of the Cold War. But much attention was given to how significantly the interaction between science and world politics has changed in the post-Cold War period.

One change has been the geographical shift in focus from East-West conflicts to other political hot spots, such as the Middle East, Southeast Asia, the former Yugoslavia and the North-South development gap.

It was pointed out, for example, that although many of the organizations created to ease East-West tensions after the Second



World War — such as NATO, the OECD and the United Nations Educational, Scientific and Cultural Organization (Unesco) — have thriving scientific programmes, networks need to be formed among scientists to address current political realities.

"What is least likely is that we'll have a repeat of the Cold War and that the institutions that came out of it will be able to take care of today's problems," said Jesse Ausubel, director of the programme for the human environment at the Rockefeller University, New York.

Several speakers commented that there is substantial scientific networking within Europe. But this remains rare in politically troubled regions.

Abdelhamid Chorfa, an economist, statistician and demographer from the Institut de Stratégie in Algiers, pointed out that the post-colonial scientific infrastructure of North Africa still lacks international networks, largely owing to political conflicts.

"Scientists still define themselves by the nation from which they come," said Eugene Skolnikoff, professor emeritus of political science at the Massachusetts Institute of Technology. **Heather McCabe**

Analysis of polio vaccine could end dispute over how AIDS originated

Could an effort to rid Africa of polio have caused the AIDS pandemic? While labs prepare to test the vaccine, researchers are at war over the theory.

Paris

Since 1992, some of the last remaining samples of an oral polio vaccine (OPV) given to millions of Africans have been locked in a fridge — itself locked in another fridge — at the Wistar Institute in Philadelphia. Only director Clayton Buck has the keys. The samples, dating from the 1950s, are suspected by some of being the origin of AIDS.

The 'OPV-HIV' hypothesis — that HIV-1 jumped to humans from contaminated polio vaccines distributed in central Africa during the 1950s — has come to prominence via *The River: A Journey Back to the Source of HIV and AIDS* (Allen Lane), a book by journalist Edward Hooper (see *Nature* 401, 325; 1999).

It is known that HIV-1 came from chimps and HIV-2 from sooty mangabeys, but researchers disagree on how the original viruses crossed the species barrier. Although Hooper's book fails to reveal a smoking gun, it offers circumstantial evidence that kidneys from these animals might have been used to produce a vaccine called CHAT (records no longer exist). In a bid to settle the matter, Buck will soon open the fridge and send samples to three laboratories to find out whether they contain any trace of infection with a precursor of HIV.

The names of the laboratories are being kept secret, says Buck, to "avoid media pressure, and to ensure that the studies are blind; so that other labs don't know who is doing what". At the same time, he says, one of the world's leading mitochondrial DNA analysis laboratories will seek to identify the primate species used to prepare the samples.

Chances of detection are slight

Buck admits, however, that the studies may not settle the matter. Only six to nine samples will be tested, which means that, even if the vaccines had been contaminated, the probability of detecting the virus is slight. He also points out that the samples are over 40 years old, and in unknown condition.

"If the results are negative, then we can say: 'Mr Hooper had an interesting idea, he did a lot of very nice documentation, but perhaps his assumptions were not valid, and we need to look elsewhere,'" says Buck, who maintains that chimps were used only to test the toxicity, not in vaccine preparation.

Results should be revealed in May, at a meeting organized by Britain's Royal Society,



In the dock: were captured African chimps used to make polio vaccine, not just to test it?

"The meeting will make the debate a little less wild than it has been in the past," says Pat Bateson, the society's biological secretary, who says opposing AIDS researchers are currently in deadlock. "The OPV-HIV hypothesis is an interesting argument, and it ought to be thrashed out." He hopes the meeting will "get agreement on what new evidence might settle the argument".

Accidents can happen. Millions of people were accidentally contaminated with simian virus 40 in the 1950s through contaminated polio and adenovirus vaccines made in monkey kidney cells; luckily it seems to have done little lasting harm. But Maurice Hilleman, director of the Merck Institute for Therapeutic Research in West Point, Pennsylvania, who carried out some of these trials, is sceptical of the OPV-HIV hypothesis. He says that chimpanzees were never used to make vaccines, as they were expensive and unsuitable for the large amounts of tissue needed.

But Simon Wain Hobson, an AIDS researcher at the Institut Pasteur in Paris, and Robin Weiss of University College, London, say Hooper's hypothesis needs to be examined. Hobson admits that the book produces no direct evidence, but says he was unable to demolish its arguments. In particular, he found Hooper's correlation in space and time between the emergence of AIDS and the vaccination campaigns "compelling".

"Why did all these strains of HIV suddenly appear at same time?" Hobson asks. He has played Sherlock Holmes in the past, showing that the AIDS virus that Robert Gallo, of the US National Cancer Institute, claimed to have discovered was identical to one earlier discovered at the Institut Pasteur, a sample of

which had been lent to Gallo. Hobson argues that the possibility that AIDS is man-made should give grounds for caution in allowing xenotransplants to proceed.

But other scientists criticize him for lending credence to what they claim is little more than a conspiracy theory. John Moore, of the Aaron Diamond AIDS Research Center in New York, fears the publicity is hampering vaccination efforts in developing countries.

Beatrice Hahn, a scientist at the Howard Hughes Medical Institute, argues that the weight of evidence suggests AIDS probably arose much earlier (see *Science* 287, 5453, 607; 2000). She believes that AIDS was transmitted to humans "via cutaneous or mucous membrane exposure to infected animal blood" through hunting, butchering or eating uncooked contaminated meat. Dirty needles could also have caused "rapid serial passage of viruses in humans, thereby facilitating viral adaptation to the new host".

Distraction from a bigger issue

Hooper, says Hahn, has "come up with a hypothesis, but failed to demonstrate that chimpanzee or sooty mangabey kidneys were even used". She argues that it is a distraction from a bigger issue, namely that there are dozens of HIV-like viruses in wild monkey populations, and that if natural transfer of AIDS viruses from chimpanzees to monkeys has already occurred, there is no reason why it should not happen again.

Hahn also points to recent mathematical models of the evolution of HIV by Bette Korber, of the Los Alamos National Laboratory in New Mexico, which calculate that the AIDS viruses jumped to humans between 1910 and 1930. Hobson argues that Korber's model relies on many assumptions, which, when combined with the margins of error, fail to kill the HIV-OPV hypothesis.

For his part, Hilary Koprowski, who developed the vaccine at the Wistar Institute, says it is ironic that he is being "demonized" in the year that the World Health Organization is making a final push to eradicate polio. "Hooper's book is based on preconceptions, not facts," says Koprowski.

He has "very mixed feelings" about the Royal Society meeting, and has yet to decide whether to attend. "This is not a response to Hooper," he says. "Hooper must come up with the proof."

Declan Butler

G. ROLLAIS, TAKEN FROM THE RIVER

Spying charges against Ukrainian researcher dropped

London The Ukrainian security services have dropped all charges against marine biologist Sergey Piontkovski who, with colleagues from the Institute of Biology of the Southern Seas in Kiev, had been under investigation since last October over his links with foreign funding agencies (see *Nature* **401**, 835; **402**, 6; 1999).

Charges against Piontkovski for illegal currency operations and 'organized crime' were due to be heard this month, but were dismissed last week, and his passport was returned. All charges were also been dropped against others involved in the case (see *Nature* **402**, 455; 1999).

Piontkovski, an expert in the dynamics of plankton populations, says he assumes the move is a "result of all that's being going on, including the reaction of the world's media and all the letters". A representative from the European aid programme INTAS — which funds Piontkovski's work and that of other Ukrainian scientists — recently visited the local prosecutor, a move that may have influenced the decision to halt prosecution.

If successful in obtaining a visa,

Piontkovski says that his immediate plan is to pursue his research at the Plymouth Marine Laboratory in the United Kingdom. He hopes to leave this week, but expresses concern that "there might be incidents at the border".

German lasers approach ever shorter wavelengths

Munich The German national research centre for particle physics (DESY) has announced a "decisive milestone" in its quest to develop an X-ray laser that materials scientists and biologists can use for the study of atomic structures at very high resolution. A research team at DESY has used a Free Electron Laser to create ultraviolet radiation with a wavelength below 100 nm, from an electron beam provided by one of DESY's linear accelerators.

The research team hopes eventually to create wavelengths as short as 6 nm, the soft X-ray range, using a 300-m Free Electron Laser currently under construction. DESY has plans for an even more powerful Free Electron Laser, to create wavelengths as short as 0.1 nm. But this depends on getting approval for its planned next-generation linear accelerator, TESLA (TeV-Energy Superconducting Linear Accelerator). A decision on TESLA is due in the next couple of years.

French plan 'virtual' nuclear tests

Paris France's Atomic Energy Commission announced a deal with the company Compaq last week to buy a 5-teraflop supercomputer for simulating nuclear tests. The machine, which can perform five trillion operations each second, will make the computer centre at Bruyères-le-Châtel, near Paris, the most powerful facility of its kind in Europe when it goes into operation in 2001.

The supercomputer forms part of the core of the French nuclear-simulation strategy that also includes the X-ray imaging facility Airix, located outside Paris, and the Laser Megajoule, being built near Bordeaux. The computer centre will also be open to scientists for other research.

Japanese officials found guilty in HIV blood case

Tokyo Judges at the Osaka district court last week recorded the first convictions in Japan's scandal around HIV-infected blood products, finding three former top officials of the Green Cross Corporation guilty of negligence. The case stems from the mid-1980s, when more than 1,800 haemophiliacs in Japan were infected with the HIV virus through blood products.

The former managers were found guilty despite evidence that there had been no clear regulatory guidance from the Ministry of Health and Welfare on the need to recall untreated blood products. The trials of a former health ministry official and a former vice-president of Teikyo University who acted as a ministerial adviser are still under way.

Internet spectrometry lab gets funding boost

San Diego Oregon State University has been awarded a \$1 million grant from the W. M. Keck Foundation to fund a plasma mass spectrometer that can be accessed over the Internet and manipulated from a distance.

Calling the spectrometer and associated equipment a 'collaboratory', university officials say that the facility, which will be completed in a few months, will permit state-of-the-art geochemical analysis of specimens from disciplines ranging from marine science to archaeology.

German green card scheme to plug IT recruitment gap

Munich German chancellor Gerhard Schröder last week announced plans to offer a European version of the US 'green card' to 30,000 foreigners, in an attempt to help plug

an estimated employment gap of 70,000 workers in Germany's booming information technology business.

Schröder says the computer experts are likely to come from Eastern European and Asian countries. Despite generally high unemployment in Germany, labour unions have signalled support for the idea, provided that the effort ends when Germany has caught up in training its own workforce. German education minister Edelgard Bulmahn has announced the creation of 40,000 training vacancies in the information technology sector this year alone.

Reed-Elsevier invests in an online future

Paris The new chief executive officer of the publishing group Reed-Elsevier, Crispin Davis, last week announced that the company is to spend £750 million (US\$1.2 billion) on reinventing itself as an Internet company in its three core divisions: science, legal and business-to-business. The group hopes that by 2002 more than 60 per cent of its scientific revenues will come from Internet projects, and that total Internet revenues will have grown to £1 billion.

According to Davis, the publishing group will focus its science effort on its online product Science Direct, and move into the

acquisition and development of software toolsets. To achieve this, it has created a £100 million venture-capital investment fund, called Greenhouse. Elsevier's profit margins on its scientific publishing activities in 1999 were 34.5 per cent, in contrast to the days when it enjoyed margins greater than 40 per cent.

Blair takes a more cautious stance on GM foods

London Britain's Prime Minister, Tony Blair, said last week that he understands "anxieties" over genetically modified (GM) foods, and admitted that such foods hold the "potential" for harm. In an article for the *Sunday Independent*, writing about an international conference being held in Edinburgh on the safety of GM foods (see *Nature* 403, 821; 2000), he wrote that the conference "recognizes the jury is still out on the application of this new technology to food and crops and there is cause for legitimate public concern".

Blair's comments have been widely interpreted as a U-turn on the Government's earlier stance, and consequently welcomed by environmentalist groups. But a government spokesperson said that Blair's comments had been misinterpreted. "We've always said we have to proceed cautiously, on the basis of science".

Crowded universities cramp more than just students' style

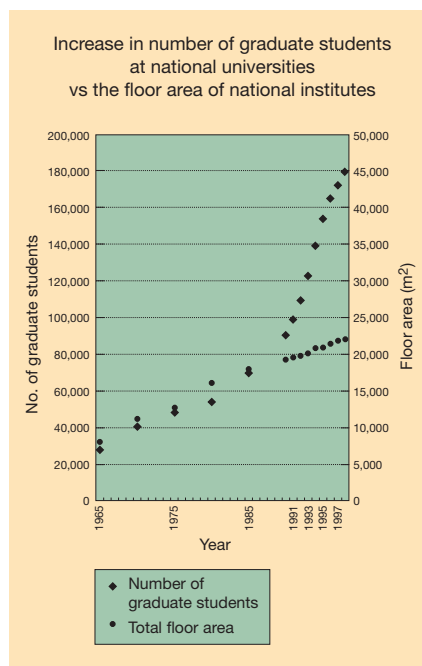
Sir— Many of the comments made by David Swinbanks (*Nature* **403**, 7; 2000) about the serious problems in Japanese universities are valid. Laboratory space in most universities is severely limited; overcrowded laboratories can be seen everywhere. Visitors from abroad are often appalled by these poor conditions and wonder why nothing is being done about them.

The main reason for the deterioration of facilities has been a lack of investment by the government in higher education and in university research over many years, even when Japan was in a period of economic growth. The recent government policy of increasing the number of graduate students without proper investment in university buildings (see figure) has aggravated the problem. University authorities who accepted more graduate students without requesting the proper investment share the responsibility for this situation.

The figure shows how the number of graduate students in Japanese national universities has increased without a concurrent increase in space. The Ministry of Education, Science, Culture and Sports (Monbusho), which is mainly responsible for these matters, expects a further increase from the current 180,000 to 250,000 in 2010. That would make the situation even worse.

Swinbanks suggests that Japanese scientists must learn to take action. But they have made efforts previously to do something about poor conditions. A survey by the Chemical Society of Japan (*Nature* **339**, 575; 1989), backed by media reports and by a strong appeal from the then president of Tokyo University, Akito Arima, forced a recognition of the miserable university research conditions in the late 1980s and early 1990s. This led to legislation in support of science and technology and to a boost in funding. However, the biggest and the most fundamental problem, that of upgrading university facilities, still remains.

At present, the Japanese government does not make public its spending priorities. It would seem sensible to provide enough research space first, and then, when the infrastructure is in place, pump in more money for the actual research. But things do not proceed that way in Japan. Arima used to be a strong advocate for the improvement of research conditions, and later became the head of both Monbusho and the Science and Technology Agency. But he was not very



Going through the roof: student numbers are rising faster than the space available for them.

successful, during his short term, in improving matters in Japan's political and bureaucratic framework. Other ministries seem to regard the problems of university facilities as matters for Monbusho alone, and they resist attempts to change the way in which public money is shared out among ministries.

The Science Council of Japan, an advisory organization for government comprising 210 members elected from various scholarly societies, made a survey of space problems and unanimously recommended that the government take urgent measures to improve the situation. But the reaction of Japan's general media has been far from enthusiastic. Most politicians are either unaware of the problem or are uninterested in it.

I hope the government will place the highest priority on construction of university buildings in its next five-year plan for Science and Technology.

Akio Yamamoto

Department of Applied Chemistry, Waseda University, 3-4-1 Ohkubo, Sinjuku-ku, Tokyo 169-8555, Japan

Patent confusion in law on new plant varieties

Sir— The news (*Nature* **403**, 3; 2000) that the European Patent Office (EPO) has reversed its decision in the Plant Genetic Systems case gives cause for concern. How and why has the Enlarged Technical Board of Appeal — which made the original decision — changed its mind?

Originally, the application for a patent on Plant Genetic Systems plants was refused because the EPO decided that the material described as 'plant cells and their resulting plants' was de facto a plant variety and was therefore unpatentable. The EPO said that the stable incorporation of foreign DNA into the genome of an existing variety merely made another variety — and thus the reconstituted plants were covered by the regulations of the International Union for the Protection of New Varieties of Plants.

The original Enlarged Board of Appeal took the most narrow definition possible of a plant variety — that is, that the incorporation of a single gene is enough to differentiate one variety from another.

Yet the EPO has now reversed its decision, based on the alleged need for an even narrower definition. This means the EPO has now defined a plant variety as being less than a single gene's separation between two plant lines.

European patent law concerning genetically modified organisms is in a mess, and needs to be completely rethought to balance the commercial, ethical and scientific aspects of this difficult question. The European Directive on Patenting Biotechnological Inventions is full of ambiguity and contradictions — so much so that two European Union states have already referred it to the European Court.

The EPO's latest decision has made it harder for biotech companies to argue their case from an environmental perspective. How will they be able to say that their products are no more risky than conventionally bred crops, while at the same time arguing that their plant and animal creations are patentable inventions and thus, by definition, new? It will not take long for organizations such as Greenpeace to start using this as an argument against genetically modified organisms.

On the same page as the news item about the EPO decision was a report that the US Patent Office has tightened up by disallowing 'speculative claims'. Even more interestingly, these new guidelines are to remain open for public comment on a website until 22 March. This should be a new model of operation for the EPO, whose decisions seem arbitrary and out of touch with the increasing public concern on the patenting issue.

How about an open consensus conference on patenting living things, organized by the EPO, as the basis for an equitable European way out of the quagmire of patents for genetically modified organisms?

John R. Porter

Royal Veterinary and Agricultural University, Agrovej 10, 2630 Taastrup, Denmark

Evolution rising from the grave

Reactivation of a dormant message signals the dawn of a new humanity.

Darwin's Radio

by Greg Bear

Del Rey: 1999. 430 pp. \$24

Michael A. Goldman

It is the dawn of the third millennium. It may be the dawn of a new humanity. It is the greatest scare since the bubonic plague. It manifests itself as cold-like symptoms in pregnant women, and as devastating malformations in their fetuses. It shakes the very sequence of the human genome. A cataclysm of biblical proportions in the making, it is dubbed Herod's influenza. A hitherto-silent human endogenous retrovirus has excised itself from the genome and become mobile and infectious once again.

SHEVA — scattered human endogenous retrovirus activation — has been seen in the past. In mass graves, where men and pregnant women had been slaughtered, their remains tested positive for SHEVA. It was also present in a well-preserved Neanderthal couple found in an ice cave — the woman's unborn child had met a violent end. The dread disease spreading across the face of the Earth today had broken out aeons before, at the dawn of *Homo sapiens sapiens*. Renegade scientists would show that SHEVA was not the awful disease Herod's influenza implied. It was, rather, Darwin's radio, conveying the message that would catapult mankind to its next step in evolution.

I am not a reader of mainstream science fiction, but for years I have told incredulous students, colleagues and friends that good, and even bad, science fiction can spark public interest in and understanding of science and its implications for our society. *Jurassic Park* has taught a generation about the possibilities and impossibilities of genetic engineering. *Gattaca* makes us think about where we're going with human genetic screening and genetic enhancement. But Greg Bear's novel *Darwin's Radio* goes one step further. It not only gets the non-scientific reader to think about details of molecular biology such as retroviruses, it challenges busy scientists to think freely about what the disjointed discoveries of the past few decades might really mean. And it reminds us just how closed to new ideas we can sometimes be. Bear restates the idea that science moves forward when orthodoxy is challenged. But the orthodoxy challenged must be last week's dogma; tampering with today's can be hazardous to one's career. This the scientists in *Darwin's Radio* learn. The pioneering thinkers of yesterday are the devoted traditionalists of today. New ideas enter science grudgingly. New paradigms are resisted with a vengeance.

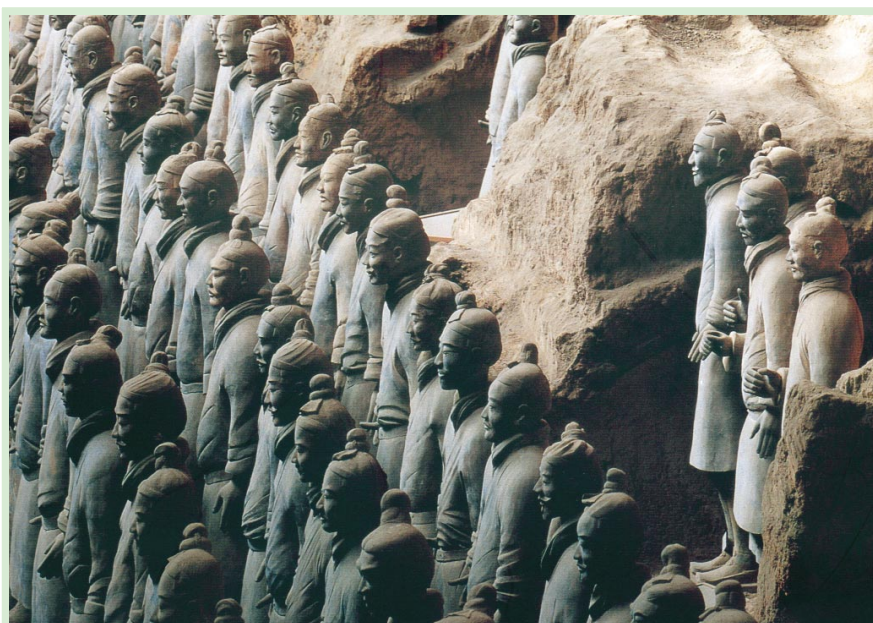
In writing *Darwin's Radio*, Bear treads on the toes of the scientific elite: "Ernst Mayr's kids are sweating ice cubes ... Dawkins is beside himself." He has mustered a cadre of facts, loosely connected and ill understood. There are little happenings at the periphery of Mendelian genetics, at the edge of neo-Darwinian theory, and what used to be beyond the realm of molecular biology. We know that bacteria contain extrachromosomal elements that transmit traits, such as drug resistance, between individuals and between species. We know that elements exist which insert in and excise from the genome, sometimes in response to environmental signals or stresses.

Such extraordinary magic isn't limited to single-celled organisms. Even before bacterial transposons, plasmids and phages were understood, Barbara McClintock had seen evidence for movable elements in maize, with all the prescience of Kaye Lang, Bear's heroine. We are faced with our growing knowledge of the human genome. It is littered with the footprints of ancient retrotransposons, ancient retroviral infections and shocking, widespread rearrangements. Stephen Jay Gould and others have advocated the notion that evolution occurs

by punctuated equilibria. The fossil record shows bursts of speciation and extinction cutting across entire fauna and flora. Yet we seldom make the connection that vertical transmission of genetic information could occur within and among higher eukaryotic species, and that clusters of transposition events could occur in response to environmental signals. Bear the science fiction writer makes the connection for us.

Most of us believe that simple, incremental changes in allele frequencies, driven by the forces of genetic drift, mutation, recombination, migration and natural selection, are enough to explain evolution — from adaptation to speciation, to the origin of higher taxa. There is no compelling evidence to the contrary, but neither is there compelling evidence in favour of the idea; we simply haven't observed or catalogued the forces and changes that create new species. Bear fills this void in our understanding with the notion that radical changes in the genome, brought about by mobilization of transposable elements such as human endogenous retroviruses, result in rapid change at the subspecies or species level.

I'm not afraid of this concept. Bear goes a little further in suggesting that such change



Buried in their thousands

The tomb of China's first emperor, the self-styled First August Emperor of Qin, is guarded by a terracotta army of up to 7,000 life-size and lifelike warriors, which were only unearthed in 1974. This is just one example of the mass

production of exquisite Chinese works of art found throughout history. *Ten Thousand Things: Module and Mass Production in Chinese Art* by Lothar Ledderose (Princeton University Press, \$60, £38) celebrates these achievements.

can occur over about a generation, an idea that might be a little too radical at the moment. However, he does mention data suggesting that fruitflies can adapt to a new environment in just a few generations of selection. He further suggests that speciation (or subspeciation) can occur in response to environmental stimuli, such as stress. This doesn't seem so incredible in view of the way in which prophages become excised from bacterial chromosomes in response to stress. Hard scientific evidence tells us that the scenario is possible. Although not every step in evolution must be preceded by a rampant retroviral infection, the possibility that such a sudden shake-up of the genome can be the fuel for evolutionary change may soon become dogma. Richard Goldschmidt's idea of the "hopeful monster" seemed strange in his day, but monsters produced by homeotic mutations are now on the cover of every textbook in genetics and developmental biology. Only the "hopeful" is missing at the moment.

One of Bear's wilder speculations is about the rapidity and directedness of evolutionary change. Mobile elements "hopping around like bugs on a hot griddle" can surely produce genetic change, but the likelihood of beneficial genetic change would be nil. Bear concedes in his novel that there isn't a simple, direct path to a new species — terrifying malformations, straight from an episode of *The X-Files*, characterize the aborted fetuses in the first three-quarters of the book.

But there is some basis for speculating that something other than a grossly random mess would result from a massive mobilization of transposable elements. First, preferred sites of integration can occur, and might ensure that a majority of elements land where they will do little harm. There is evidence that some repetitive-sequence elements in the human genome harbour specific regulatory elements, like retinoic-acid response elements, which respond to developmental signals. Thus, it is possible that simple insertion of a retroviral element in the vicinity of a gene could result in an alteration in the timing or positioning of its expression in early development. The result could be a different, yet perfectly viable, organism.

As an established science fiction writer who takes pride in keeping ahead of scientific developments, Bear exploits the latest in nanotechnology in his grand scheme of evolution. It's hard for me to imagine Bear's "mighty Wizard in our genes", the consummate computer in our genomes, responding to input from the environment, carrying out crosstalk with other individuals and species, and playing back several different scenarios for ushering the human species to new levels of achievement. But computer scientists say that the next generation of data processors will depend not on incremental improvements in the silicon microchips of today, but on a quantum leap into the realm of

molecules whose sequence of chemical bonds encodes information on a nanometric scale. If we can use the DNA molecule to carry out calculations in ways that were unimaginable just ten years ago, is it impossible to believe that nature has used the same molecules to encode instructions about instructions we do not yet understand?

Darwin's Radio, no matter how preposterous or prophetic one thinks the science, is superb 'hard' science fiction, speculating about the connections among well-known facts. It is not a textbook of conventional ideas. It is an entertaining, even riveting story, delivered poetically. It portrays scientists as real people, responding to the intense politics of the biomedical world, the funding imperative in public and private sector alike, and the terrifying challenge of a disease that threatens to decimate the human species. It takes a hard look at the challenges faced by a woman scientist with radical ideas, and the excitement of discovering a totally new way of looking at biological evolution. Whether you read it to pass a cold, snowy night by the fire, or to free your mind for the new paradigms that will emerge in the next millennium, I promise you an engaging journey. ■

Michael A. Goldman is in the Department of Biology, San Francisco State University, San Francisco, California 94132-1722, USA.

.....

Around the world via Greenwich

Globes at Greenwich: A Catalogue of the Globes and Armillary Spheres at the National Maritime Museum

edited by Elly Decker

Oxford University Press: 1999. 592 pp.

£95, \$160

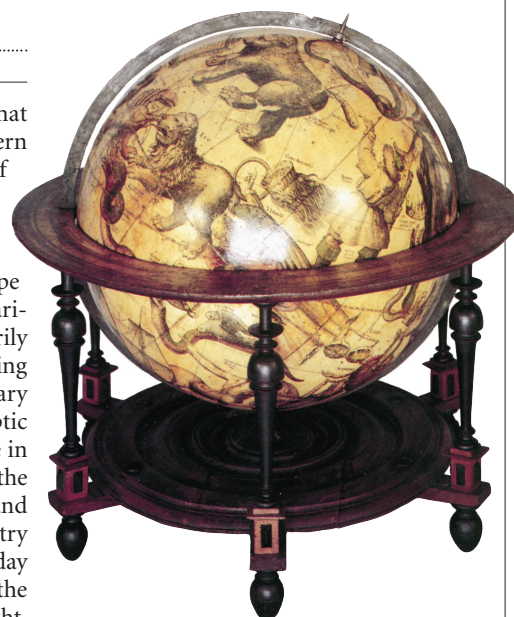
J. L. Heilbron

Like many other things, globes are not what they used to be. The ordinary, modern terrestrial variety has but one degree of freedom — a rotation around an axis inclined to the vertical — as a hint that the Earth belongs to the Solar System. Early modern Western globes, the dominant type in the collection of Britain's National Maritime Museum at Greenwich, customarily had two degrees of freedom, to allow a setting in latitude as well as longitude, and auxiliary circles representing the horizon, the ecliptic and the colures. Often these globes came in pairs, one celestial and one terrestrial, for the easier learning not only of constellations and land masses, but also of spherical geometry and elementary astronomy. If set for day and place, a pair of globes can indicate the stars visible above the horizon at night, the directions of sunrise and sunset, the

length of the day, the Sun's altitude at noon, and so on. As we read in the *Encyclopédie*, the great inventory of the knowledge of the Enlightenment, globes were once "mathematical instruments".

A typical standard globe of the seventeenth and eighteenth centuries had a diameter of 32.5 centimetres. The largest specimen in the National Maritime Museum's collection, which dates from 1688, has a diameter of 108 cm and thus an area equal to that of some 60 sheets of A4 paper. A considerable amount of information and error can be inscribed on such a sphere. Its maker, Vincenzo Coronelli, took advantage of both possibilities, filling his oceans and other blank spaces with accounts of discoveries, depictions of peoples, ships and sea creatures, and definitions of celestial and terrestrial coordinates. At the other end of the globular scale were balls 7.5 centimetres in diameter, introduced in 1654 by Joseph Moxon. These usually came in boxes the inside of which bore a depiction of the constellations. Gentlemen carried them in their pockets to orientate and amuse themselves during their travels.

Most of the museum's globes were printed on paper gores (slips shaped to fit together into a sphere). In her catalogue, which occupies three-quarters of *Globes at Greenwich*, Elly Decker gives the arrangement of the gores (where relevant) and many other construction details of the globes and armillary spheres (representations of the principal circles of spherical astronomy) owned by the museum. She also provides significant geographical and cartographical information about each item and transcriptions of their inscriptions as well as customary data about maker, dimensions, materials, date and provenance. All this is done to the very highest standards. Silke Ackermann's



Gone global: a 1613 representation of the Earth.

description of the dozen Islamic globes in the collection attains the same level.

The Islamic objects differ from the Western ones in several ways. They depict only the heavens, most are made of metal, and all present the constellations as mirror images of the Western depictions. This last peculiarity arises from the fact that we see asterisms from within the sphere of the stars and their simulacra from outside the globe. In the West, an asterism consisting of, say, a catalogue facing the viewer and holding an instrument in her right hand would be rendered on the globe from behind, the instrument still in her right hand, whereas on Islamic globes she would face the viewer and hold the instrument in her left hand. Western globe-makers solved the problem of representing the constellations by interchanging back and front, their Islamic counterparts by transposing left and right. Thus, the fecund globes contain the seeds not only of geography and astronomy, but also of comparative cultural psychology and the sociology of scientific knowledge.

The third of the book not given over to the catalogue consists of nine essays, four of them by Dekker. The other authors include Jonathan Betts (on clockwork globes), Anne Leane (construction and conservation) and Karen Lippincott (globes in art). From Leane we learn that a standard globe with printed gores was formed of papier mâché moulded over a ball of metal or wood. The moulded spherical shell was then cut in half, provided with an internal axis and perhaps additional braces, pasted back together, and coated with plaster before the gores, a protective layer and a final varnish were applied. Lippincott discusses the deployment of globes primarily as upbeat icons denoting power and order. The best and fullest of the essays, like the catalogue itself, is the work of Dekker. It describes in detail the "uncommonly handsome globes" in the collection: the oldest (1537), the biggest (1688), the most important pair (by Mercator, 1541 and 1551), the finest pair (by John Senex, around 1730), a pair with brass gores, a silver terrestrial globe with the heavens engraved on its inside, a model of the Moon, and so on, down to a dozen English miniatures.

Within its terms of reference, which exclude social history, *Globes at Greenwich* is a superb work — with one grievous exception. Its very many figures, well over 100 of them in colour, are of a most uneven quality and clarity — some are excellent, others just acceptable, many others muddled, and a few disgraceful. Easily the worst are the black and white illustrations in Lippincott's essay on globes in art. These blemishes spoil the book for readers interested in globes more as objects of beauty and utility than as items in an inventory.

J. L. Heilbron is at Worcester College, Oxford
OX1 2HB, UK.



Water culture: Lake Texcoco and the ancient Aztec capital of Tenochtitlán, site of today's Mexico City.

A window on scientific progress

Science and Technology in World History: An Introduction

by James E. McClellan III & Harold Dorn
Johns Hopkins University Press: 1999. 404 pp.
\$18.95, £13 (pbk)

Gunhard Æ. Oravas

For some time there has been an urgent need for an introductory textbook that would afford a panoramic critical view of the development of technology and science in the light of their mutual influences throughout history. This historical account achieves its basic aim of demonstrating that, with the exception of quite recent history, technology has always influenced science, not the other way round.

The book encompasses all the great ancient civilizations. Interestingly, the authors devote considerable attention to the technological hydraulic cultures of the New World: the Maya, Aztec and Inca cultures. European history of technology and science is surveyed from medieval and Renaissance times, through the Industrial Revolution to the present day. For example, the theory of evolution of Charles Darwin and Alfred Russel Wallace is traced through to the decoding of the molecular structure of DNA. Also included are Germany's reform of its universities and the establishment of technical high schools. The latter were devoted to both teaching and original research, in an effort to professionalize practical science, in areas such as chemical technology, electrotechnology and precision optics.

But the text is peppered with mistakes. The law of free fall was first established by

the Spanish Dominican theologian and political philosopher Domingo Francisco de Soto, in his lectures on Aristotelian physics at the University of Salamanca. These were published first in 1545, then more completely in 1551. De Soto — and not, as the book says, Galileo — was the first to recognize that free fall is a uniformly accelerated motion. He was also the first to derive the law of free fall by using the geometric form of proof of the Mertonian mean-speed rule.

Galileo claimed these discoveries as his own while being fully aware of the brilliant medieval work that proved fundamental to the new science of mathematical mechanics. Experiments on the free fall of heavy bodies had been carried out by Ioannes Philoponos of Alexandria in about 517, and by Giovanni Benedetti and by Simon Stevin during the early Renaissance, and all three experimenters had refuted Aristotle's law of free fall. Galileo obtained his idea of inertial motion from Benedetti's work and from reports of the lectures at the Collegio Romano. However, under Platonic influence he conceived kinematic inertial motion as being 'natural' circular motion, contrary to Benedetti's concept of rectilinear inertial motion. Thomas Harriot, one of the founders of modern algebra and a correspondent of Johannes Kepler, discovered sunspots and the satellites of Jupiter many months before Galileo.

Kepler, unlike Galileo, was a man of genius with an incredibly rich imagination. He believed in the principle of analogy as the very basis of invention and demonstrated this repeatedly in his work. Kepler's work on geometric optics was full of new ideas and insights, but is not included in this book. This is a grave omission, since it represented a vigorous beginning to modern geometric optics, even containing the

first suggestion of the wave theory of light, the foundation for all later work on optics. Kepler's mathematical work was tremendously important in the invention of calculus, for he was the first mathematician since Archimedes to make completely free use of infinitesimals. In contrast, Galileo's mathematics was unoriginal, frequently clumsy and quite obsolete by the time it was published.

To highlight another inaccuracy in the book, Robert Hooke and Robert Boyle *did not* invent the vacuum pump; they only improved on Otto von Guericke's vacuum pump after Boyle received information on Guericke's invention in 1657. Hooke was the first scientist to set down the fundamental problem of astronomy with his four principles in 1674. And in 1678 he proposed the inverse square law of universal gravitation — first advanced by Kepler in 1604 — which Isaac Newton used in his famous work *Principia* on celestial mechanics.

Numerous omissions at times produce an unsatisfactory text. The mathematician Christiaan Huygens, for example, deserves far more attention than he receives. He was the only mathematician whose research had Archimedean rigour, and he excelled in many areas. Unlike Newton and Gottfried Leibniz, Huygens was also a theoretical and experimental physicist, a theoretical and practical astronomer, a technologist, an engineer and a skilled portrait artist. Furthermore, he not only used infinitesimal analysis and improved the method of maxima and minima, but also employed the coordinate method of Pierre Fermat and René Descartes in his mathematical research.

In addition to his outstanding mathematical work on the pendulum, rectification problems and the theory of probability based on the concept of expectation, Huygens established his conservation of work principle. He applied this to the fracture of light beams, and it is still used today in the limit analysis of simple beams. He designed and built the first accurate cycloidal pendulum clock, the first spring-driven watch fitted with a balance-wheel, and a high-resolution refracting telescope incorporating his micrometer invented in 1658 and his compound eye-piece (the Huygens ocular), which minimized chromatic aberration and which is still used today. He also developed the undulatory theory of light which, in modified form, remains part of modern physics.

However, in spite of these frustrations (and others I have not listed), this highly accessible book is much needed and should be well received. ■

Gunhard Æ. Oravas is in the Department of Civil Engineering, McMaster University, Hamilton, Ontario L8S 4L8, Canada.

Science in culture

Entomology meets art

A Consilience, a video installation at the Natural History Museum, London.

John Whitfield

Put bluntly, the video installation *A Consilience* consists of five entomologists at London's Natural History Museum talking about their work with the Belgian artist Jan Fabre, great-grandson of the eminent entomologist Jean-Henri Fabre. The project is the latest instalment in the museum's arts programme, produced in collaboration with the Arts Catalyst, an agency that works to bring scientists and artists together (see <http://www.artscatalyst.org/htm/new.htm>).

The end product, though, is less a documentary than a dream the museum might have after the visitors have gone home for the night. The entomologists are dressed as representatives of the insect groups they study, while Fabre plays a beetle, a theme of his work. With their wings and goggles, they look like the eccentric inventors of ill-fated early flying machines.

The researchers take to their roles like naturals, and you feel that this might actually be how they spend their evenings. Their performances are informed by their knowledge — Dick Vane-Wright, Keeper of Entomology, dresses as what he describes as a “universal butterfly” (pictured), and enacts the behaviour of “skippers, swallowtails, cabbage whites and so on”. They also mirror the impressions a lay person might have about the different insect groups: the flies and wasps come over as more manic than the butterfly and beetles. The cast is pleased with the outcome: “*A Consilience* has a real arthropod buzz to it,” says Martin Brendell, who both curates the museum's beetle collection and plays one in the video. “Jan's video editing gets across the diverse activity patterns of an insect very well,” adds (fly, and sometime head of veterinary entomology) Martin Hall.

The arthropod buzz emerges from style rather than content: images jump and stutter, making it difficult to see the details of the costumes. People and insects shape-shift into one another — who is playing what emerges only after lengthy viewing. The dialogue is hard to follow; only isolated words and phrases emerge from the conversations. So, instead of leaving with some new information, the viewer has to work out a meaning for themselves.

The other star of the piece is the museum itself; Fabre shows it as a mystical, enchanted site. Instead of enlightenment we get Gothic — a cheeky subversion of the usual stuff of a museum visit. Most of the film shows the entomologists prowling through the bowels of the museum's collections, down spookily lit corridors piled high with bottled and stuffed beasts. Sometimes they bump into Fabre, and stop for a handshake and a chat.

Some scenes could have come from a horror film — one of the best sequences shows the human/insects emerging from a silk cocoon — and there are some overtly ritualistic props, such as an orb and a crucifix. The Natural History Museum is often compared to a cathedral: *A Consilience* suggests that the priests of taxonomy are down in the crypt conducting a black mass. Wandering around the other galleries, it is hard to shake the feeling that the whole museum is, in fact, a bizarre installation.

But as well as this, the work shows scientists breaking through the thick crust of detail, routine and struggle that overlies a life in research and getting to the real reason they do what they do — because the world is a source of exciting, beautiful and weird things, and not just grants. And it shows that scientific creativity, like its artistic equivalent, is not predictable or well-behaved. Inspired by working with Fabre, Vane-Wright has since written an article for the arts journal *Janus* (April, 39–43; 2000) on the combinatorial patterns of butterfly wings: “What is the relationship between the 50,000–100,000 or so butterfly patterns that exist, and the billions that could exist? Are they a random subset of the possible and, if not, in what ways are they constrained?” He also took the opportunity to speculate on “possible analogies between biological diversity, art and aesthetics”.

“Normally, you have to bottle such ideas up for some dotty book written in old age, when everyone knows you are past it,” he says. “This way I might get to enjoy the answer while I'm still alive.” ■

A Consilience finished its run at the Natural History Museum, London, recently.

John Whitfield is a subeditor at *Nature*.



Reversing Rorschach

Ink-blot tests might be able to tell us more about creativity than personality.

Richard Gregory

The Swiss psychiatrist Hermann Rorschach (1884–1922) died only a year after the publication of his famous ink-blot test. Making ink blots to elicit weird and wonderful perceptions was a children's game — *Klecksographie* — that he probably knew well, for he was nicknamed “Klex” as a child. Rorschach produced only one major publication, *Psychodiagnostics: A Diagnostic Test Based on Perception* (Huber, Berne, 1921), but it launched just about the largest literature in psychology.

He found that some ink blots were better than others at eliciting strange perceptions that might be used clinically: “In the first place, the forms must be relatively simple; complicated pictures make the acquisitions of the factors of the experiment too difficult. Furthermore, the distribution of the blots on the plate must fulfil certain requirements of composition or they will not be suggestive, with the result that many subjects will reject them as ‘simply an ink blot’ without consideration of other possible interpretations.”

He used symmetrical patterns: “It has a disadvantage in that it tends to make answers somewhat stereotyped. On the other hand, symmetry makes the conditions the same for right- and left-handed subjects; furthermore, it facilitates interpretation in certain inhibited and blocked subjects. Finally, symmetry makes possible the interpretation of whole scenes.”

Some plates evoked responses to details, others to the whole pattern. Colours evoked kinaesthetic movement responses. Some blots were ‘harder’ than others, and for some ‘whole’ answers were almost impossible. He classified answers according to whether the subject saw a fixed form, movement or colour. The type of response was seen as important for psychological assessment and diagnosing mental illnesses.

Most significant were the reported forms, seen most clearly by “pedants and depressives” (academics?): “Most interpretations are determined by the form of the blot alone, both in normal and abnormal subjects. The subject searches among his visual memories for that one which in form, especially in outline, most closely resembles the entire figure or one of its details. In accomplishing this, he does not visualize the object ‘seen’ as moving, but as a fixed form.” Animals were seen most frequently, and seeing oneself was most common in schizophrenics.

Rorschach used just ten ink blots. This launched a major industry in clinical psy-

chology that has only recently begun to decline. I am not competent to assess its reliability for assessing individuals; but I suggest that reversing the test — from kinds of people to kinds of patterns — might show what stimulates creativity. This is a clear experimental question: which kinds of pattern evoke the richest variety of perceptions and ideas?

This Reversed Rorschach should reveal principles of creativity. For a start, one may think of realistic pictures as representing external objects, whereas ink blots and abstract paintings evoke internal creations. Which patterns or pictures are most evocative should tell us what switches us on most powerfully to create new perceptions and ideas. It should reveal the creative nature of mind, for generating perceptions and conceptions for art and perhaps also science.

The refrigerator in my kitchen sports ‘magnetic poetry’. This has words or phrases as units which can be moved around. These are far more evocative than individual letters. Perhaps much the same applies to concepts for teaching science: too ‘atomic’ facts will not suggest connections or concepts; too ‘molecular’ will inhibit originality. It seems that there is an optimal degree of structure for the seeds of creativity to grow into perceptions and ideas. Teaching should start with optimal molecules rather than atoms. This mental chemistry of creation might be developed by experiments with Reversed Rorschach ink blots, and perhaps generalized to language and thinking.

What of creation by painters? Do their partially completed paintings change before their eyes? Do phantasms — to be nurtured



A scowling, mustachioed man? A walrus?
A chinese dragon mask?

or exorcised — suggest the next move of the brush? Perhaps much of this applies to writers and scientists; for as ink-blot fantasies show, conceptual and perceptual processes are entwined in an intimate dance.

This may be too close for comfort to the predicament of train drivers: it is reported that, in Britain, about two trains pass a red signal every day. The driver is looking for long periods at muddled, quite symmetrical structures — rather like staring at an unfolding ink blot. Does the scene change form and meaning in his sight? Can red lights merge or disappear?

It seems to be profoundly true that all perceptions are loosely controlled hallucinations. Thus colours, triggered by wavelengths of light, are created by brain processes and psychologically projected into the external world. No doubt there is something out there, though as science advances how this is described departs ever more from any appearances. Whatever is out there, perception's mental worlds can kill and maim when crucially mismatched to physical reality.

Although their clinical validity may be dubious, ink blots might evoke creativity in controlled ways, allowing us to understand and perhaps tame the danger lurking behind our fanciful eyes.

Richard Gregory is in the Department of Experimental Psychology, University of Bristol, 8 Woodland Road, Bristol BS8 1TN, UK.

Which patterns are most evocative should tell us what switches us on to create new ideas.

Danger — hard hack area

“Sequence your genome at home, and set science free!” cry the biopunks.

Paul McAuley

Many people predicted that VirCon 2010, the first open meeting of the biopunk movement, would end in a riot. In truth, it was as privately exciting and as publicly dull as any science conference. From their besieged underground culture, the clandestine surfers of the new wave in biology are emerging blinking into the daylight and, dare one say, into respectability.

But VirCon 2010, held in a dilapidated midtown New York hotel, was not without friction. Despite the rule that no biological material could be brought in, there was a ruthless but futile inspection by officers of the Food and Drug Agency. Several people suspected of being undercover federal agents or snoops from biotech companies were summarily ejected, and the press was barred, which led to strange scenes outside the hotel as TV journalists were videoed typing into a laptop to communicate with conference delegates just inside the lobby.

I was allowed to cover the event from the inside because of personal contacts made while covering the pursuit and arrest of Kevin ‘Freaky-Deaky’ Miles, the man who claimed to have turned the Amazon rainforest luminescent — and because I’m a science fiction writer, and biopunks love SF.

The delegates were mostly young, white males under 25, dressed in everything from baggies and T-shirts, through goth black and multiple piercings, to business suits. All had self-inflicted gene hacks: feathers or scales instead of hair; bands of chromatophores on their foreheads; motile tattoos. And of course, unlike the pasty-faced, overweight cliché of computer hackers, the biopunks were bursting with health, their skin and eyesight perfect, their muscle definition superb, their energy seemingly boundless.

That energy was needed: the convention operated on a 24-hour basis, and no one slept. The biopunks took over the hotel, talking over laptops displaying DNA source codes; poking through half-dismantled sequencers; and attending workshops on resetting metabolic clocks, overwriting junk DNA, perturbing quantum effects in microtubules, and, of course, meme viruses.

In many ways it was like a science meeting in the so-called Golden Age, before research was taken over by big business and commercial confidentiality strangled the exchange of ideas. Biopunks are, of course, scornful of secrecy. They are pathological braggarts; their culture is based on the idea of open



source, from pirated genomes placed on servers from Cuba to Finland, to hacks disseminated in philes and e-zines with titles like *Triplet Threat* and *They’re Made Out Of Meat*. The only difference is that their published work stands not by peer review but by utility, and that almost everyone uses their hacker handle rather than their real name.

One major exception to the last was the grey eminence of biopunk, Professor Jack Lovegrove, who became a legend after he was fired from the University of Kansas for teaching a practical course in evolution and jailed for distributing pirate DVDs containing the sequence of the entire human genome.

“Garage science is the wave of the future for biology,” he told me in a brief interview. “The big companies are tied down by class-action suits and vicious regulations imposed

by frightened, short-sighted politicians. But anyone can isolate DNA in their kitchen, and sequence it rapidly with cheap, modern RNA chips. Twenty years ago, DNA had to be cut by expensive enzymes and sequenced one base pair at a time. Now, the sequencers found in any physician’s surgery can be modified to run in parallel, so copyrighting the human genome is a joke. Anyone here could sequence it in a couple of days, starting with one of their own cells.”

Lovegrove’s keynote address, “Post-humanism and how to achieve it”, was packed, but he was at pains to point out that most of his ideas were at the ‘golden vaporware’ stage. “We’re interested in radical ideas, not frost-proofing strawberries, but the rumours that these kids have found the cure for cancer, the key to immortality or boosting intelligence are a crock. Anyone who’d done any of those things would have boasted about it by now, set up his own company, or sold out to Cytex or Dow.

“Of course, people are trying out radical techniques on themselves, but medicine has a long history of self-experimentation. This is no secret society of hyperintelligent supermen — just idealistic kids who love science and believe that information should be free, not locked up by copyright and litigation. All our ideas are discussed openly, although not everyone will be able to understand them.”

As for the idea that biopunks are using meme viruses to modify human behaviour, Lovegrove was scathing. “The meme thing is mostly over. Sure, there were pranks. People had fun injecting their victims with pseudo-religious visions of Elvis or Princess Di. But now the drug cartels have muscled into memes, the craze has died down. Besides, changing belief systems is a hard hack. I’d say that the virulent campaigns against genetic modification died out from natural causes, not from infection by some imaginary supermeme.”

He’s right. There are limits to what kids in home-made labs can do. But in the heady atmosphere of the convention, with biopunks partying to the limit or deep in jargon-ridden conversation, their tattoos and chromatophore bands flashing and blushing, it’s hard to believe that some of their more fantastic claims won’t remain science fiction. ■

Paul McAuley was a biology lecturer at the University of St Andrews until he became a full-time writer in 1996. His latest novel is *Shrine of Stars* (Gollancz).

JACEY

Of ice and elephants

Daniel P. Schrag

What caused the great ice ages? Accurate dating of the end of one of these ice ages makes us reconsider which part of the climate system drives severe glaciation.

For over a century, geologists have turned to astronomy for explanations of the cause of ice ages during the Pleistocene epoch (the past 1.6 million years). In the 1870s, James Croll¹ first suggested that ice ages were caused by changes in the amount of solar irradiance received at the poles as a result of changes in the shape (with a frequency of about 100,000 years), tilt and wobble (with frequencies of about 40,000 and 20,000 years) of the Earth's orbit around the Sun. Despite the success of this astronomical theory, some fundamental questions remain. We still don't know how subtle changes in the pattern of solar irradiance are amplified to produce such spectacular changes in climate — cooling the deep ocean almost to the freezing point and extending ice sheets thousands of kilometres towards the Equator. And we don't understand why ice ages occur in both hemispheres simultaneously when the changes in solar irradiance from orbital variations have opposite effects in the north and south.

Many theories for the origin of ice ages have centred on the Northern Hemisphere as the connection between orbital variations and climate. In the 1930s, Milankovitch² suggested that orbitally induced variations in solar irradiance at 60° N drove the waxing and waning of ice sheets in North America and Europe. This was a reasonable hypothesis as most of the volume of ice during the glacial maximum was located in the Northern Hemisphere. Recently, the 'ocean conveyor belt' in the North Atlantic has been proposed as an important amplifier of climate variation, magnifying subtle temperature and precipitation changes near Greenland. But is the North Atlantic, or even the Northern Hemisphere, really running the show?

One reason for concentrating on the North Atlantic is that many of the highest-quality, best-dated records originate from this region. But, like the story of the wise men and the elephant, could we be seeing only part of the climate system? The story tells of an emperor who sends his wise men to investigate a gift from a foreign land, an exotic animal called an elephant. But the wise men have grown blind from reading their books, and after each feels a different part of the strange beast, they return to the emperor in total confusion.

On page 61 of this issue, Henderson and Slowey³ provide an accurate date for the end

of the second-to-last ice age, calling into question the Northern Hemisphere connection between orbital variations and climate. If we think of the climate as an elephant, Henderson and Slowey's finding is a reminder of how vast the elephant is, or at least it should make us re-examine which end might be the head.

Since the 1950s, the ratio of oxygen isotopes in marine microfossils has been the standard measure of Pleistocene climate change as it reflects, in part, the size of continental ice sheets. But precise dating of the

oxygen isotope records has not been possible beyond the range of radiocarbon dating (30,000 years or so), making it difficult to compare these records with astronomical data. Imbrie and colleagues⁴ proposed that complex ice-sheet dynamics, paced by orbital variations, forced the ending of ice ages, and they created a timescale (SPECMAP) for the oxygen isotope record based on this theory (Fig. 1). This timescale predicts that the end of the second-to-last ice age occurred 127,000 years ago.

Henderson and Slowey used an improved method of uranium–thorium dating, which allows accurate dating of much older sediments, to show that the midpoint of the end of this ice age was much older, at $135,000 \pm 2,000$ years ago. To achieve this level of precision the authors overcame many technical obstacles. This work represents a huge improvement in the absolute dating of the oxygen isotope record, and may redeem previous estimates of an even earlier date for this deglaciation⁵. But the real significance of this result is their claim that the new date is consistent with deglaciation driven by orbital variations in solar irradiance, either in the Southern Hemisphere or in the tropics, but not in the Northern Hemisphere.

Henderson and Slowey are not alone in suggesting that the Southern Hemisphere may drive the ice-age cycles. Data from the Vostok ice core — near the centre of Antarctica — shows a warming at the end of the last glacial maximum that appeared 1,000 to 2,500 years before the first major warming in Greenland⁶. Climate models have suggested a possible connection between orbital variations and changes in sea ice around Antarctica⁷, including a link to atmospheric carbon dioxide⁸. But the story from Antarctica has been complicated by results from an ice core from Taylor Dome, close to the Pacific margin of Antarctica, which has a history more like Greenland than Vostok⁹. And we still don't know how climatic change in the Southern Hemisphere could cause ice sheets in North America to melt.

So what about the tropics? Some wise men (and wise women) who study the Pacific argue that the vast amount of warm water in the Pacific Ocean is the dominant force affecting global climate, and that the Pacific is relatively insensitive to climate in other parts of the world¹⁰. Last year, Clement and colleagues¹¹ suggested that ice-age cycles

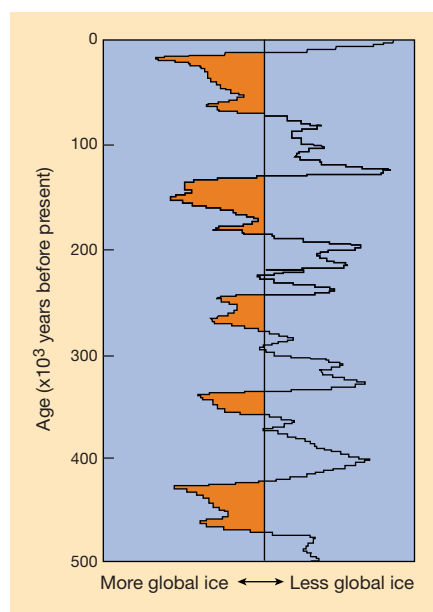


Figure 1 Timescale of ice ages during the late Pleistocene. The Spectral Mapping Project (SPECMAP) record is based on oxygen isotope data from marine fossils that reflect the size of continental ice sheets. The record for the past 500,000 years indicates that ten major ice ages and warm periods occurred during this time. Together these make up five ice-age cycles, which occur at essentially the same frequency (every 100,000 years or so) as changes in the shape of the Earth's orbit around the Sun. The Earth has been experiencing a warm period (or deglaciation) for the past 10,000 years or so. Henderson and Slowey³ now provide an accurate date for the midpoint of the previous deglaciation (135,000 years ago). This is significantly different from that predicted by SPECMAP (127,000 years ago), suggesting that we may need to rethink our picture of what drives the ice ages.

might be related to the direct effects of orbital cycles on the tropical Pacific, through changes in the frequency and intensity of El Niño. Unfortunately, appropriate sites in the tropical Pacific are difficult to find, so high-resolution, well-dated records that rival those from Greenland or the North Atlantic do not yet exist.

Henderson and Slowey's dating of the second-to-last deglaciation allows us to match climate changes and orbital cycles with new confidence, and forces us to think about how the many parts of the climate system interact. If the link between orbital and ice-age cycles does lie in the Southern Hemisphere, then we must explain how changes in the south are communicated to the north, perhaps by way of the tropics. An interesting idea proposed by Broecker¹² suggests a seesaw relationship between north and south, whereby Greenland and Antarctica are out of step during rapid climate changes (over timescales of less than 1,000 years). Although it is not yet clear how such rapid climate changes are related to the larger

cycles of ice ages, if at all, such ideas are a step in the right direction. Progress in this area depends on obtaining well-dated records from the tropical Pacific, as well as resolving the spatial and temporal pattern of climate changes over Antarctica. Without direct observations of the whole climate system, we may forever argue, like the wise men, about the exact nature of this climate beast. ■

Daniel P. Schrag is at the Laboratory for Geochemical Oceanography, Department of Earth and Planetary Sciences, Harvard University, 20 Oxford Street, Cambridge, Massachusetts 02138, USA.

e-mail: schrag@eps.harvard.edu

1. Croll, J. *Climate and Time* (Appleton, New York, 1875).
2. Milankovitch, M. in *Handbuch der Klimatologie* (eds Koppen, W. & Geiger, R.) 1–176 (Gebrüder Borntraeger, Berlin, 1930).
3. Henderson, G. M. & Slowey, N. C. *Nature* **404**, 61–66 (2000).
4. Imbrie, J. *et al.* *Paleoceanography* **7**, 701–738 (1992).
5. Winograd, I. J. *et al.* *Science* **258**, 255–260 (1992).
6. Blunier, T. *et al.* *Nature* **394**, 739–743 (1998).
7. Kim, S.-J., Crowley, T. J. & Stossel, A. *Science* **280**, 728–730 (1998).
8. Stephens, B. B. & Keeling, R. F. *Nature* (in the press).
9. Steig, E. J. *et al.* *Science* **282**, 92–95 (1998).
10. Cane, M. A. *Science* **282**, 59–61 (1998).
11. Clement, A., Seager, R. & Cane, M. A. *Paleoceanography* **14**, 441–456 (1999).
12. Broecker, W. S. *Paleoceanography* **13**, 119–121 (1998).

Cancer

p53 sends nucleotides to repair DNA

Guillermina Lozano and Stephen J. Elledge

The p53 protein actively suppresses tumour formation; when it is mutated, the road is open for the development of several forms of cancer. Mutations in p53 or the pathway that directly regulates it have been found in over 80% of human tumours, so there has been intense research into the p53 gene and its product. Several roles for the p53 protein have been identified, to which Tanaka *et al.*¹ (on page 42 of this issue) now add another. It is one, moreover, that may offer a fresh strategy for antitumour drug development.

The initial discovery that p53 is a tightly regulated transcription factor that is activated in response to DNA damage provided a biological basis for understanding its involvement in maintaining the integrity of the genome². One of p53's functions in the damage response is the activation of genes that initiate apoptosis — programmed cell death³. Selective removal of severely damaged cells is thought to protect an organism from cancer. p53 has also been shown to arrest the cell cycle in response to DNA damage, thus preventing the replication of damaged DNA⁴. Some experiments suggest that cells arrested in the G1 phase of the cell cycle by p53 remain arrested permanently, so preventing damaged cells from ever proliferating⁵.

Both of these mechanisms prevent cells, once damaged, from contributing to tumorigenesis. But does p53 contribute to damage

repair in the first place? From the work of Tanaka *et al.*¹, the answer seems to be yes. The authors have identified a new p53-regulated gene, *p53R2*, which turns out to encode a subunit of the enzyme ribonucleotide reductase (RNR).

Ribonucleotide reductase catalyses the rate-limiting step in the production of deoxyribonucleotide triphosphates (dNTPs) required for DNA replication and repair⁶. It is an $\alpha_2\beta_2$ tetramer composed of two dissimilar subunits. The large subunit, R1, contains the allosteric regulatory sites that maintain and balance dNTP pools; the small subunit, R2, contains a binuclear iron centre and a tyrosyl free radical that is essential for the enzymatic conversion of ribonucleotides to deoxyribonucleotides. Mammalian ribonucleotide reductase is regulated by the cell cycle⁷. The R2 subunit is made in the late G1 phase before DNA replication, and disappears in late S or early G2. In contrast, the R1 subunit is produced throughout the cell cycle. Paradoxically, ribonucleotide reductase is in the cell cytoplasm, presumably producing dNTPs that diffuse into the nucleus for DNA replication⁸.

The product of the *p53R2* gene is an R2 subunit that was identified by screening a p53-inducible colon-cancer cell line for differentially expressed genes. p53 binds a DNA sequence in the first intron of *p53R2* that is required for directly activating its transcription. The p53R2 protein is highly related to the normal R2 subunit, but it differs in one crucial aspect: it is found in the nucleus (Fig. 1).

This change in venue led Tanaka *et al.* to speculate that replication requires the previously known cytoplasmically synthesized source of dNTPs, whereas repair needs concentrated nuclear sources near the sites of damage. To test whether *p53R2* has a role in repair, the investigators inhibited it with antisense DNA and observed a sharp decrease in incorporation of dNTPs into

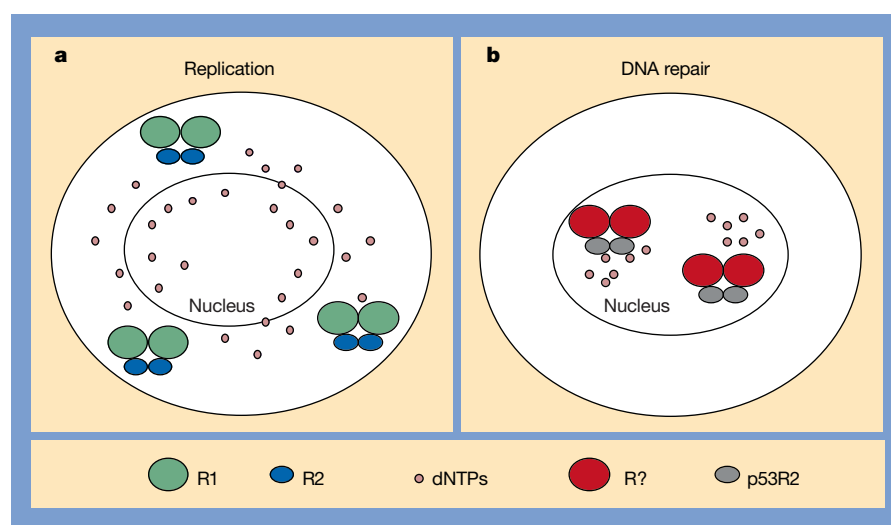


Figure 1 Differential composition and localization of ribonucleotide reductase (RNR) during DNA replication and repair. a, During the S phase in the cell cycle, the cytoplasmically localized RNR causes the production of deoxyribonucleotide triphosphates (dNTPs) in the cytoplasm which diffuse into the nucleus for DNA replication. R1 and R2 are the two subunits of the RNR tetramer. b, Tanaka *et al.*¹ show that, in the presence of DNA damage, the *p53R2* gene is induced and, together with an unknown large subunit protein, its product causes dNTP production in the nucleus to facilitate DNA repair. A cell in the G1 or G2 phase is shown.

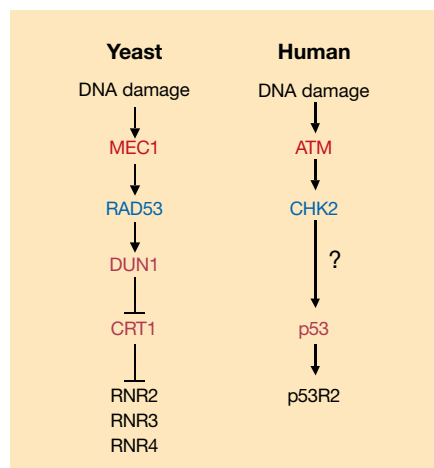


Figure 2 Similarities in the DNA damage-response pathway in yeast and mammals. In yeast, the protein kinase DUN1 is activated. Subsequently DUN1 phosphorylates and inhibits the ability of the CRT1 repressor to bind to the promoters of several RNR (ribonucleotide reductase) genes¹². (There is also a weaker CRT1-independent induction pathway, but it is not as well characterized.) The activation of DUN1 is controlled by MEC1 (mammalian homologue, ATM/ATR) and RAD53 (mammalian homologue, CHK2).

DNA after it had been damaged with ultraviolet light or adriamycin. They also showed that cells in the G1 and G2 phases of the cell cycle, which do not normally make R2, induce a DNA-repair capacity in a p53-dependent manner, suggesting that p53R2 may be most important for repair in phases of the cell cycle other than the S phase.

Finally, Tanaka *et al.* demonstrate that cells that fail to make p53R2 are more sensitive to killing by DNA-damaging agents. These overall results could be strengthened by analysis of cells containing deletions of p53R2. But even as they stand, the data provide the first direct evidence of a role for p53 in DNA repair and maintenance of genomic stability.

It has, however, long been appreciated that dNTPs are involved in DNA repair. Inhibition of ribonucleotide reductase is known to sensitize cells to DNA damage, and the transcriptional induction of RNR-encoding genes occurs in response to DNA damage^{9,10}. In the yeast *Saccharomyces cerevisiae*, four genes encode ribonucleotide reductase and each is inducible by DNA damage. The DNA damage-response pathway (also called the checkpoint pathway) that controls induction of RNR genes is largely conserved with that in humans which regulates p53. Yeast have no p53 gene. Instead they have a circuit in which, in response to DNA damage, the protein kinase DUN1 is activated. The details of the pathway are shown in Fig. 2, along with the mammalian homologues of the yeast proteins involved. That this pathway is conserved from yeast to humans

underscores the importance of dNTP regulation in cell survival and validates the relevance of the study of model organisms to human disease.

These new results¹ raise several questions, the most important of which concerns the role of p53R2 in tumorigenesis. None of p53's other transcriptional targets has been found to be mutated in tumours. Perhaps this is because p53 activates so many genes that eliminating just one causes only a slight increase in tumorigenic potential compared with eliminating them all at once. But p53R2 may be different. It is directly involved in the process of repair, and other components of DNA-repair pathways are known to be mutated in cancers. For example, mutations in genes concerned with nucleotide excision repair cause skin cancer, and mutations in mismatch repair genes cause colon cancer¹¹. So a repair gene such as p53R2 may turn out to be a more important target of p53 for tumour suppression.

One goal in understanding the role of p53 in tumorigenesis is to uncover a weakness in p53-mutant tumours that can be exploited to selectively kill them. Might the lack of p53R2 induction provide such an opportunity? Because cells lacking p53R2 are sensitive to DNA damage, perhaps what residual resistance they have is due to dNTPs produced by the cytoplasmic form of ribonucleotide reductase. If one could identify inhibitors specific for the cytoplasmic form of the enzyme, these inhibitors might selectively sensitize p53-mutant cells to DNA-damaging chemotherapeutic agents. Specifically inhibiting ribonucleotide reductase

activity in the cytoplasm might allow repair synthesis to proceed in wild-type cells through p53R2 expression, while completely shutting off synthesis in p53-mutant tumours which would then lack functional R2 and p53R2 subunits.

To identify such inhibitors it will be necessary to find out which subunits of ribonucleotide reductase combine with p53R2 to activate dNTP synthesis in the nucleus. Will it be R1 or an unknown R1-related subunit? It would be fitting if the most basic of biological processes, dNTP synthesis, proved to be the Achilles heel of p53-mutant tumours — and most satisfying. ■

Guillermo Lozano is in the Department of Molecular Genetics, University of Texas, M. D. Anderson Cancer Center, Houston, Texas 77030, USA.

e-mail: gglozano@notes.mdacc.tmc.edu

Stephen J. Elledge is in the Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, Texas 77030, USA.

e-mail: selledge@bcm.tmc.edu

1. Tanaka, H. *et al.* *Nature* **404**, 42–49 (2000).
2. Kastan, M. B., Onyekwere, O., Sidransky, D., Vogelstein, B. & Craig, R. W. *Cancer Res.* **51**, 6304–6311 (1991).
3. Gottlieb, T. M. & Oren, M. *Semin. Cancer Biol.* **8**, 359–368 (1998).
4. Ko, L. J. & Prives, C. *Genes Dev.* **10**, 1054–1072 (1996).
5. Linke, S. P., Clarkin, K. C. & Wahl, G. M. *Cancer Res.* **57**, 1171–1179 (1997).
6. Jordan, A. & Reichard, P. *Annu. Rev. Biochem.* **67**, 71–98 (1998).
7. Bjorklund, S., Skog, S., Tribukait, B. & Thelander, L. *Biochemistry* **29**, 5452–5458 (1990).
8. Engstrom, Y. *et al.* *EMBO J.* **3**, 863–867 (1984).
9. Elledge, S., Zhou, Z. & Allen, J. *Trends Biochem. Sci.* **17**, 119–123 (1992).
10. Filatov, D., Bjorklund, S., Johansson, E. & Thelander, L. *J. Biol. Chem.* **271**, 23698–23704 (1996).
11. Lengauer, C., Kinzler, K. W. & Vogelstein, B. *Nature* **396**, 643–649 (1998).
12. Huang, M., Zhou, Z. & Elledge, S. J. *Cell* **94**, 595–605 (1998).

Circadian clocks

Heartfelt enlightenment

Ueli Schibler

Most animals and plants adapt their activity to daily cycles of light and dark. However, their behavioural and physiological cycles are not merely dictated by changes in light intensity, but are controlled by an internal timing system called the circadian clock. As indicated by its name — *circa diem* means 'about a day' — the circadian clock can measure time only approximately and must be readjusted every day by light. So a circadian timing system needs a light-detector and a way to transfer the information it detects to the molecular clockwork¹.

How photons entrain biological clocks is not completely understood, but the paper by Whitmore *et al.*² on page 87 of this issue describes a startling finding: that circadian oscillators in zebrafish hearts and kidneys can be synchronized by light without any

input from the eyes. Previously, the same group showed³ that the expression of *Clock*, a master regulatory gene of animal circadian systems, undergoes daily cycles in many zebrafish tissues. These include not only organs known to contain circadian oscillators, such as the pineal gland and the eye, but also peripheral organs, such as heart, kidney and spleen. Rhythmic cycles of *Clock* gene expression persisted for several days in cultured zebrafish organs, showing that there are self-sustaining circadian oscillators in these tissues. Similar observations have also been made in the fruitfly *Drosophila* and in mammals, in which most body cells are believed to contain circadian clocks^{4,5}.

Whitmore *et al.*² noted that the rhythmic changes in the amounts of *Clock* messenger RNA in cultured hearts and kidneys were greater under a light–dark cycle than in

NATURE

100 YEARS AGO

A report on the commercial value of the metric system, with special reference to the classification of German iron manufactures, was recently forwarded to the Wolverhampton Chamber of Commerce by the British Consul at Amsterdam, and is referred to in the *Board of Trade Journal*. The Consul states that the iron and steel manufacturers' unions of Germany have adopted a uniform system of dimensions for articles of universal consumption at home and abroad. Angle iron of all descriptions, flanged boiler ends or fronts for Cornish or Lancashire boilers, the boilers themselves, and iron and steel tubes and all fittings connected with them, such as valves, cocks, T pieces, are made, so far as flange, diameter, and working lengths are concerned, in normal standard sizes, in order that every part of one work may be procured at once to fit every corresponding part of another construction... At present a committee of the engineers' associations is occupied in endeavouring to fix a metric thread for bolts and screws, nuts, bolt-heads, &c., as the present universal normal standard (the Whitworth) is so differently constructed by different works that the parts are not as interchangeable as should be the case. These classifications are naturally making more and more progress in Germany, not in the iron trade alone, but in other manufactures. In view of these facts, The Consul points out that Germany and the Continent generally will have a constantly increasing advantage over British manufactures in the future in foreign countries, unless the metric system be fully and entirely adopted by Great Britain. From *Nature* 1 March 1900.

50 YEARS AGO

Mr. Douglas Kennedy is the director of the English Folk Dance and Song Society, and as such knows his subject. He has written a delightful little book of some 158 pages with twenty illustrations and two distribution maps. Dancing throughout the ages has been for humanity a psychological release and otherwise. Shortly after the first use of mustard gas in the First World War, an American hospital unit took over a complete British tented hospital in Rouen. The strain of the personnel was very great; but it was completely relieved by the introduction of dances. Dancing was the channel for release of the overwrought emotion. From *Nature* 4 March 1950.

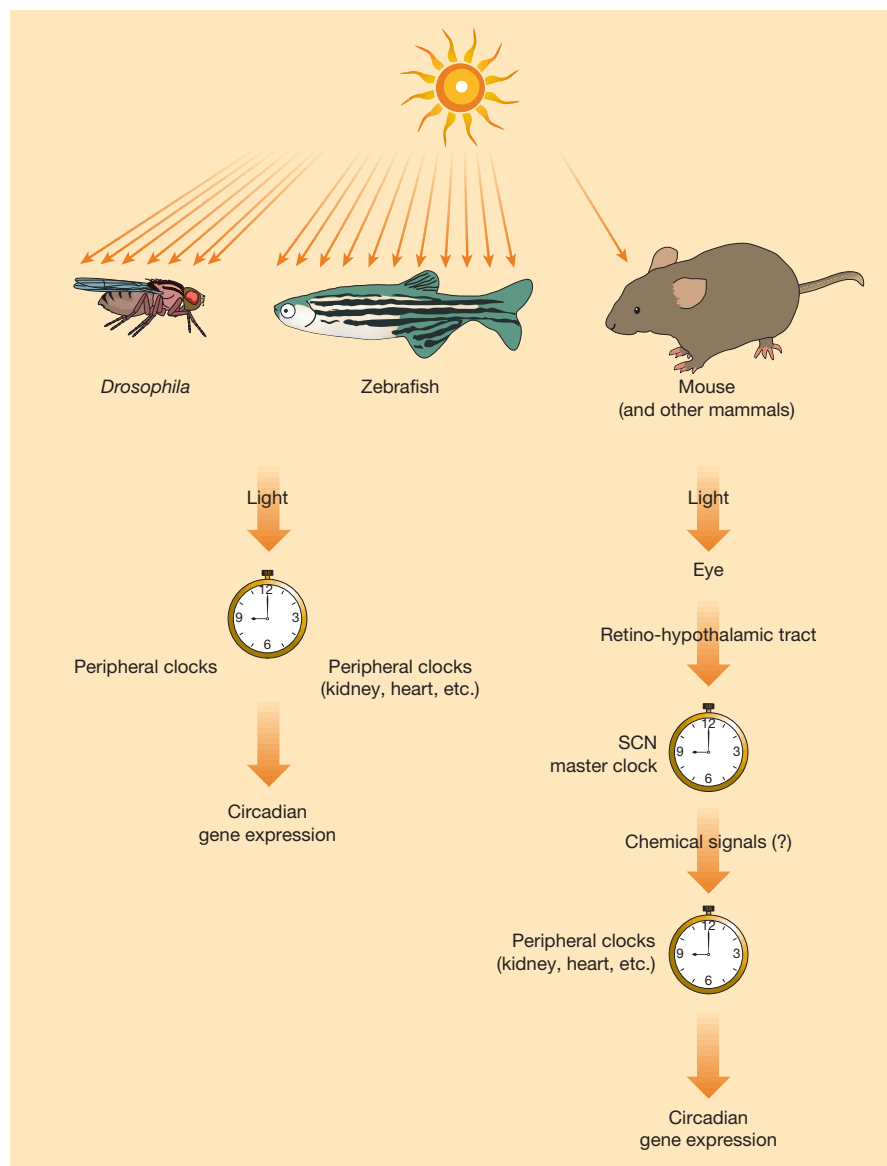


Figure 1 Synchronization of circadian clocks in cells outside the nervous system. Circadian oscillators have been found in peripheral cells, such as heart and kidney cells, of insects (*Drosophila*), fish (zebrafish) and mammals (rat and mouse). Whitmore *et al.*² have now shown that in zebrafish, as in *Drosophila*, these oscillators can be synchronized directly by light. In mammals, light entrains master oscillators in neurons of the suprachiasmatic nucleus (SCN) in the brain, which then synchronize oscillators in peripheral cells using chemical signals. Like mammals, *Drosophila* and zebrafish contain brain structures that are important for circadian behaviour and physiology (lateral pacemaker neurons in *Drosophila*, pineal gland in zebrafish). Whether these specialized brain structures contribute to the entrainment of molecular oscillators in peripheral *Drosophila* and zebrafish cells is unknown.

constant darkness. They therefore considered that daily light cycles seem to entrain the oscillations, preventing them from dampening over time. To test this possibility, the authors monitored *Clock* gene expression in heart and kidney organ cultures kept in reversed light–dark cycles. Two days after the reversal of the light–dark regimen, the phase of the *Clock* messenger RNA cycle had been offset by 12 hours, showing that heart and kidney clocks could indeed be entrained by light.

Even embryonic zebrafish cell lines can synchronize to light–dark cycles. When these cells were kept in constant darkness,

Clock messenger RNA levels stayed the same throughout the day. But exposure to two light–dark cycles caused a significant daily fluctuation of these transcripts, and the rhythmic *Clock* messenger RNA cycle then persisted for two days under constant darkness.

Molecular circadian pacemakers consist of genes specifying proteins that can activate or repress gene transcription, and whose abundance or activity (or both) can be fine tuned by other regulatory proteins (such as protein kinases). The transcriptional activators enhance the expression of the repressors until the latter reach a concentration

sufficient to turn off transcription of their own genes. The repressor proteins have a limited lifespan so they eventually decay, and a new wave of repressor gene transcription commences. The properties of the system have evolved to generate oscillations with a period of about 24 hours^{6,7}. All known components of animal circadian clocks have homologues in both *Drosophila* and mouse¹. These include the transcriptional activator Clock, the putative transcriptional repressors Timeless and Period (three isoforms in the mouse: mPer1, mPer2 and mPer3), and blue-light receptors called cryptochromes (two isoforms in mouse: Cry1 and Cry2)^{8–11}.

The function of cryptochromes in animal circadian clocks is enigmatic. In plants and *Drosophila*, cryptochromes clearly participate in the light entrainment of circadian clocks, but in the mouse their function as light receptors is still controversial^{8,10,12}. Nonetheless, mouse cryptochromes are uncontested components of the mammalian pacemaker, as mutant mice with defects in both cryptochrome genes, *cry1* and *cry2*, lack circadian rhythms when kept in constant darkness and display constitutive expression of the genes *mper1* and *mper2* (refs 8,10,13). The photoreceptors used to entrain peripheral zebrafish clocks are unknown, but Whitmore *et al.* may be able to use their *in vitro* system to test whether cryptochromes are valid candidates. The recording of action spectra may be the first step towards this goal.

In mammals, peripheral clocks are synchronized by different mechanisms from those in *Drosophila* and zebrafish (Fig. 1). *Drosophila* and zebrafish are both small and semitransparent, and their peripheral oscillators can be directly light-entrained by

light receptors^{2,5,9,11} (such as cryptochromes in *Drosophila*). In mammals, which are opaque, this mechanism is obviously not feasible for setting the time in peripheral oscillators. In these organisms, light adjusts the circadian pacemaker in part of the brain called the suprachiasmatic nucleus through signals from the eyes, and this master pacemaker then synchronizes peripheral clocks using chemical signals¹⁴. Curiously, neither retinal rods nor cones are required for the light entrainment of mammalian clocks, so the light-capturing cells and their photoreceptors remain to be identified^{15,16}.

Ueli Schibler is in the Département de Biologie Moléculaire, Sciences II, Université de Genève, 30, Quai Ernest Ansermet, Genève CH-1211, Switzerland.

e-mail: ueli.schibler@molbio.unige.ch

1. Dunlap, J. C. *Cell* **96**, 271–290 (1999).
2. Whitmore, D., Foulkes, N. S. & Sassone-Corsi, P. *Nature* **404**, 87–91 (2000).
3. Whitmore, D., Foulkes, N. S., Strahle, U. & Sassone-Corsi, P. *Nature Neurosci.* **1**, 701–707 (1998).
4. Balsalobre, A., Damiola, F. & Schibler, U. *Cell* **93**, 929–937 (1998).
5. Plautz, J. D., Kaneko, M., Hall, J. C. & Kay, S. A. *Science* **278**, 1632–1635 (1997).
6. Rosbash, M. *et al.* *Cold Spring Harb. Symp. Quant. Biol.* **61**, 265–278 (1996).
7. Young, M. W., Wager-Smith, K., Voshall, L., Saez, L. & Myers, M. P. *Cold Spring Harb. Symp. Quant. Biol.* **61**, 279–284 (1996).
8. Vitaterna, M. H. *et al.* *Proc. Natl Acad. Sci. USA* **96**, 12114–12119 (1999).
9. Stanewsky, R. *et al.* *Cell* **95**, 681–692 (1998).
10. Okamura, H. *et al.* *Science* **286**, 2531–2534 (1999).
11. Emery, P., So, W. V., Kaneko, M., Hall, J. C. & Rosbash, M. *Cell* **95**, 669–679 (1998).
12. Cashmore, A. R., Jarillo, J. A., Wu, Y. J. & Liu, D. *Science* **284**, 760–765 (1999).
13. van der Horst, G. T. *et al.* *Nature* **398**, 627–630 (1999).
14. Brown, S. A. & Schibler, U. *Curr. Opin. Genet. Dev.* **9**, 588–594 (1999).
15. Freedman, M. S. *et al.* *Science* **284**, 502–504 (1999).
16. Lucas, R. J. & Foster, R. G. *Endocrinology* **140**, 1520–1524 (1999).

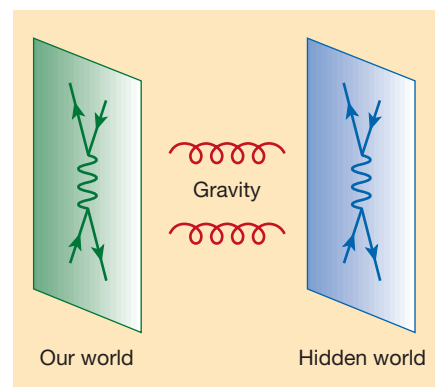


Figure 1 In ‘brane-world’ scenarios it is proposed that our world is a three-brane propagating within an ambient space–time of higher dimension. Particle physics is confined to the brane and interacts with the extra dimensions and other ‘hidden’ three-branes through gravitational interactions. Randall and Sundrum^{1,2} argue that it is possible to have an extra dimension of infinite extent if the ambient space–time is curved in a particular way.

Sundrum papers were in fact suggested by developments in string theory. String theory was originally conceived of as a theory of oscillating objects with one spatial dimension, or strings. It is now known that in order for string theory to be mathematically consistent it must also include a rich spectrum of extended objects of higher dimension: ‘membranes’ with two spatial dimensions, ‘three-branes’ with three, and so on. Understanding the properties of these branes is very much an active area of research.

One of the key ideas in the papers by Randall and Sundrum is that the four-dimensional space–time we observe at everyday scales is actually the evolution in time of a three-brane moving through an ambient space–time of higher dimension (Fig. 1). In such ‘brane-world’ scenarios, particle physics is confined to the brane but the particles can interact with the ambient space–time through gravitational interactions. When all the extra spatial dimensions of the ambient space–time are compact, these interactions can be so weak as to have escaped detection by experiments thus far.

Before the paper by Randall and Sundrum¹ it seemed almost obvious that the ambient space–time in which the three-brane lives should not be of infinite extent. In the simplest example with a single, flat, infinite extra dimension, Newton’s law of gravity would be modified: the force between two masses on the three-brane would be inversely proportional to the cube, not the square, of the distance between them. The surprising observation made by Randall and Sundrum is that if the extra dimension were infinite, and if the embedding space–time has a very particular curvature, then it is possible for the predicted

Physics

Brane new worlds

Jerome Gauntlett

It is common sense that at everyday scales we live in a world with three large spatial dimensions. Lisa Randall and Raman Sundrum have recently made the bold suggestion in *Physical Review Letters*¹ that an extra dimension of infinite extent may supplement the three spatial dimensions we observe. Not only do they claim that this can be entirely consistent with current observations, they also argue in a second paper that closely related scenarios could hold the key to some fundamental issues in attempts to unify particle physics with gravity².

The idea that the four space–time dimensions that we observe (three space and one time) could be supplemented by extra dimensions was first put forward by

Theodor Kaluza and subsequently developed by Oskar Klein in the 1920s. The motivation then was to achieve a unification of Maxwell’s theory of electromagnetism and general relativity, which is Einstein’s theory of gravity. They considered an extra fifth dimension curled up in a small circle that would be undetected at much longer length scales. Although Kaluza and Klein’s model turns out to give wrong predictions and so cannot be a correct description of the real world, the beautiful idea of extra dimensions with finite extent (‘compact dimensions’) has become a central component of string theory — the leading candidate to unify all of the known forces of nature.

Some of the ingredients in the Randall–

corrections to Newton's law to be consistent with experimental results.

This remarkable insight has stimulated a flurry of subsequent papers developing the ideas and determining the implications for cosmology and particle physics. Directions being pursued include verifying in more detail how four-dimensional general relativity emerges on the three-brane, and studying ways in which the framework can be embedded in string theory.

The implications for particle physics are particularly exciting because other closely related ideas published by Randall and Sundrum² may provide a resolution to the 'hierarchy problem', one of the most important issues in going beyond the Standard Model of particle physics. The Standard Model is a fantastically successful theory of three of the forces of nature: electromagnetism, the weak nuclear force and the strong nuclear force. It unifies electromagnetism and the weak nuclear force into the electroweak force at characteristic length scales of around 10^{-17} cm (roughly the limit of the scales probed by current accelerators). There are strong arguments that unifying particle physics with the fourth known force, gravity, into a theory of quantum gravity (string theory, say) will have a characteristic length scale of around 10^{-33} cm. It is very difficult to account for such a vast gap between these two scales without fine tuning the theoretical parameters to an extraordinary extent. This is the hierarchy problem.

The conventional approach for dealing with this problem is to invoke a new symmetry between matter and forces called supersymmetry, which could be detected by the next generation of particle accelerators. An alternative approach (which might also include supersymmetry) is to assume that our world is a three-brane. The first proposals^{3,4} along these lines considered a single three-brane embedded in a space-time with at least two extra dimensions that are compact and flat. These dimensions could be as large as 1 mm without violating known experiments. In a string theory setting it is possible that the string length scale could be just below that probed by current accelerators, rather than about 10^{15} times smaller as previously supposed. These schemes are fascinating although they do introduce another hierarchy that needs explaining.

By contrast, Randall and Sundrum suggest that a curved ambient space-time with one extra dimension might provide a better setting. They consider a slab of this space-time bounded at each end by a three-brane. (Slabs of space-time bounded by branes were first introduced in string theory in ref. 5.) One of these branes is our world and the other is a 'hidden' world. Particles on our three-brane interact with the extra dimension and the hidden world through

gravitational interactions. By assuming that the distance between the branes is very small, Randall and Sundrum showed that these interactions are weak enough to be consistent with experiment. They also showed how the hierarchy of scales on our three-brane can be accounted for in a fascinating way by the curvature of space-time without introducing any extra hierarchy.

The Randall-Sundrum papers do not provide a detailed model of particle physics beyond the Standard Model. Indeed there are considerable difficulties to be overcome to achieve this goal. But they have provided exciting alternatives to conventional ways in which people thought the unification of particle physics and gravity might occur. Particle physicists, string theorists and cos-

mologists are currently devoting much effort to developing ideas and deriving predictions that could be tested by the next generation of particle accelerators and gravity experiments. If Randall and Sundrum are on the right track, there could be exciting experimental evidence in the near future. ■

Jerome Gauntlett is in the Department of Physics, Queen Mary and Westfield College, Mile End Road, London E1 4NS, UK.

e-mail: j.p.gauntlett@qmw.ac.uk

1. Randall, L. & Sundrum, R. *Phys. Rev. Lett.* **83**, 4690–4693 (1999).
2. Randall, L. & Sundrum, R. *Phys. Rev. Lett.* **83**, 3370–3373 (1999).
3. Arkani-Hamed, N., Dimopoulos, S. & Dvali, G. *Phys. Lett. B* **429**, 263–272 (1998).
4. Antoniadis, I., Arkani-Hamed, N., Dimopoulos, S. & Dvali, G. *Phys. Lett. B* **436**, 257–263 (1998).
5. Horava, P. & Witten, E. *Nucl. Phys. B* **475**, 94–114 (1996).

Cognitive neuroscience

Imaging in the fourth dimension

Thomas Elbert and Andreas Keil

The 1990s, the decade of the brain, saw enormous developments in neuroimaging. Structural details of the brain can now be reconstructed non-invasively as three-dimensional images; and small, task-related changes in cerebral blood flow, even in the deepest recesses of the brain, can be seen. The principles of the functional organization of the brain are being uncovered, and it seems that not only initial processing stages but also complex aspects of perception and cognition can be mapped onto brain

structures^{1,2}. Nevertheless, cognitive neuroscientists find themselves in an odd, but perhaps not surprising, situation. As the number of studies increases, so does the number of conflicting results. On page 80 of this issue, Patel and Balaban³ provide an example of what has been missing in neuroimaging — a new approach that adds time as the fourth dimension.

Usually, three-dimensional images of cerebral blood flow, metabolic changes or the activity of populations of neurons are

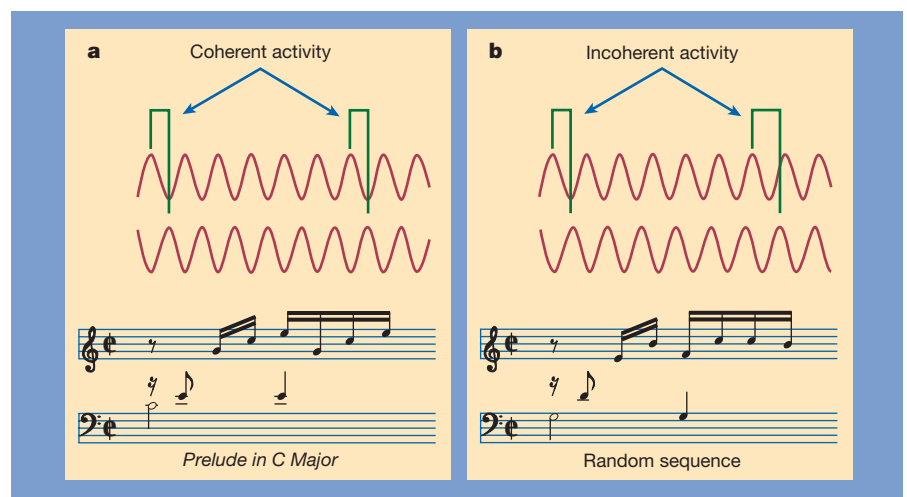


Figure 1 Music and random notes produce different neural responses. Oscillatory brain activity evoked by repetitive auditory stimulation shows different temporal relationships between widely separated recording sites, depending on stimulus properties. a, Patel and Balaban³ found that predictable, melody-like tone sequences are associated with coherent activity and more constant phase lags of oscillatory responses recorded from distant channels. This synchronization of brain activity in distant areas (the traces shown above the music) may reflect perceptual integration. The example shown here, Bach's *Prelude in C Major*, should produce high coherence. b, In contrast, a random sequence of tones showing identical rhythmic structure but no melody should produce incoherent activity of distant channels, with less synchronization between brain regions.

acquired during a given task or perceptual activity (while the subject is listening to Bach's *Prelude in C Major*, for example). The temporally static image can account only for a static cerebral response. But the brain's response is actually dynamic and self-organizing over time. So the perception of music and its neural code must be reflected in the neural dynamics — in both space and time. It is not only important to identify the neurons, neuronal assemblies or brain regions that respond to a given input. We must also develop techniques that allow the systematic classification of the temporal dynamics underlying elements of information processing. But adding time as the fourth dimension to three-dimensional space is not easy.

Patel and Balaban³ study brain dynamics using stimulus-related magnetoencephalographic responses. Subjects hear different sequences of tones that are switched on and off in rapid, 40-Hz sequences. Examination of the stimulus-related brain response that cycles at 40 Hz, the so-called steady-state response, allowed Patel and Balaban to determine how the timing of neural responses varied with different tone sequences. They found a relationship between the phase (but not the signal power) of the steady-state response and the frequency of the acoustic stimulus. That is, the timing of the neuronal response depended on the properties of the stimulus.

Intriguingly, these phase fluctuations vary with the structures of the tone sequences. Between-site phase coherence, which indicates synchronized activity between brain areas, was most pronounced for tone sequences that resembled melodies. Generalizing this outcome, the *Prelude in C Major* should produce higher inter-channel phase coherence than the same tones shuffled in random sequence (Fig. 1). Would Bach do better than the Beatles? We do not know, but now we can study how limited brain regions track the changes in pitch of auditory sequences as a piece of music is played.

Steady-state responses are a valuable tool for monitoring activity in different sensory modalities. To exploit this approach fully, we will need to understand how such brain responses are produced. In general, if the interval between successive stimuli is short enough, the transient evoked response to one stimulus will not have died away before the next stimulus is delivered. The compound response that appears is the steady-state response. There are various ways in which transient responses can sum over time to produce a steady-state response, and these fall into two groups. For a linear system, transient and steady-state descriptions of the system's behaviour are equivalent, and a simple superposition of transient evoked responses with the appropriate time lags

should fully predict the steady-state responses⁴. However, neural assemblies are nonlinear elements. If a nonlinear system is stimulated periodically, harmonics, combination frequencies and subharmonic components may evolve⁵.

Neither of these simple principles of organization accounts for the observation by Balaban and Patel that the phase of the steady-state response follows the pitch of the auditory stimulus more strongly for scales than for melodies. Obviously, higher areas of brain influence the auditory cortex and related structures by 'top-down' processes, tuning their responses according to contextual cues and previous learning. So coupled oscillations between higher-order and sensory cortices may explain why what sounds like noise to an adult is music to the ears of a teenager.

Attempts to segregate brain function into distinct modules are limited because the nervous system tends to operate through the intercommunication of task-relevant subsystems. So the simplistic modular approach needs to be complemented by modelling, in space and time, the network that incorporates the different modules. In the visual system, several different types of discrimination can be processed in the same small

area of cortex. A similar phenomenon has been seen in the motor system⁶. Likewise, a single type of stimulation of two digits can produce two opposite, use-dependent effects on the spatial relationship of the cortical representations of the digits, depending on the nature of the discrimination condition used⁷. In other words, multiple maps, specific to different modes of discrimination or tasks, share the same region of cortex. So the three-dimensional modular approach provides us with seemingly conflicting results: there are many three-dimensional shadows in a four-dimensional world. It is time to add time!

Thomas Elbert and Andreas Keil are in the Department of Psychology, University of Konstanz, D-78457 Konstanz, Germany.
e-mails: thomas.elbert@uni-konstanz.de
andreas.keil@uni-konstanz.de

1. Frith, C. D. & Friston, K. J. in *Cognitive Neuroscience* (ed. Rugg, M. D.) 169–195 (MIT Press, Cambridge, Massachusetts, 1997).
2. Elbert, T. in *Magnetism in Medicine* (eds Andr , W. & Nowak, H.) 190–262 (Wiley, New York, 1998).
3. Patel, A. & Balaban, E. *Nature* **404**, 80–84 (2000).
4. Gutschalk, A. et al. *Clin. Neurophysiol.* **110**, 856–868 (1999).
5. Pantev, C., Roberts, L. E., Elbert, T., Ross, B. & Wienbruch, C. *Hearing Res.* **101**, 62–74 (1996).
6. Karni, A. et al. *Nature* **377**, 155–158 (1995).
7. Braun, C., Schweizer, R., Elbert, T., Birbaumer, N. & Taub, E. *J. Neurosci.* **20**, 446–450 (2000).

Animal biology

Beauty is ova-rated

Female ducks are choosy when it comes to mating. Some male mallards (*Anas platyrhynchos*) are much more attractive to females than others, and females that mate with these 'preferred' males seem to raise more chicks to adulthood. It has generally been assumed that this bias reflects a genetic advantage conferred by the father.

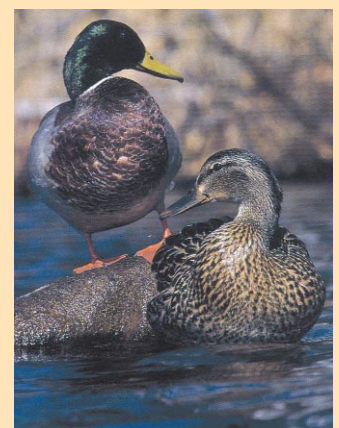
Elsewhere in this issue (*Nature* **404**, 74–77; 2000), Emma Cunningham and Andrew Russell propose a different explanation. They have compared the clutches of eggs produced by females after mating with more attractive males with those produced by females mated with less attractive males. They find that pairings with preferred males result in bigger eggs being laid.

Chicks that hatch from large eggs are more likely to survive the critical first few days after hatching. So the higher viability of the offspring

of attractive males may have nothing to do with the genetic legacy of the father — instead, it may result from increased maternal investment in the eggs. When Cunningham and Russell controlled for egg size, the attractiveness of the father made no difference to the health of the chicks.

So why do female ducks invest more energy in eggs that are fathered by more attractive males? It is likely that male attractiveness is linked to some characteristic that makes their offspring more successful in the long run. Females would then invest more in these eggs to further their own breeding success.

Male mallards are not good fathers — females do all the work of rearing chicks. So males cannot be selected by females for their paternal qualities. But they do defend feeding areas around their mates during the breeding



season, so attractiveness may be linked to the ability of a male to fend off other mallards. This ties in with the fact that females prefer males from early-hatching clutches, who tend to be bigger and stronger.

Cunningham and Russell point out that researchers need to be careful, when studying the influence of male attractiveness or dominance on the viability of offspring, to allow for the effects of differential maternal investment rather than attributing all differences to paternal genetics.

Rachel Smyly

Planetary science

Brimstone on Mars

Harry Y. McSween Jr

We normally think of the Red Planet as having a veneer of rust, but its ruddy surface is also abnormally rich in sulphur. Where did all this sulphur come from, and what might it tell us about sulphur cycling between different geochemical reservoirs on Mars? On page 50 of this issue, Farquhar and co-workers¹ provide the first unambiguous evidence that sulphur in martian rocks once resided in the atmosphere.

The Viking and Mars Pathfinder landers analysed the chemistry of the martian soil and found high concentrations of sulphur² (Fig. 1). Because of the high oxidation state inferred for the rusty surface, and spectroscopic evidence for the presence of sulphate, the sulphur in the soil is thought to occur mostly in the form of sulphates. No martian soil samples are available for analysis on Earth, but so-called SNC meteorites, which are thought to be igneous rocks from Mars³, can be analysed. Until now, only the sulphides in these meteorites had been studied^{4,5}.

Sulphur is an unusual light element in that it has four stable isotopes and can exist in several oxidation states. Geological and biological processes separate isotopes according to their masses, and these 'fractionations' are often most pronounced when the oxidation state is changed. For example, sulphate-respiring bacteria on Earth reduce sulphur and, in the process, produce relatively huge fractionations between ³²S and ³⁴S (ref. 6).

Farquhar *et al.*¹ have separated both oxidized and reduced sulphur from SNC meteorites and analysed the isotopic composition. The results are surprising: sulphur isotopes have been fractionated independently of their masses. Mass-independent fractionation is known to occur during only a few processes, and it seems that the only mechanism that could have been applicable to these meteorites is chemical reactions in the atmosphere⁷.

To substantiate this surmise, Farquhar *et al.* performed laboratory experiments to quantify the isotopic effects of gas-phase reactions on the main sulphur species thought to be present in the martian atmosphere. The distinctive fingerprint of atmospheric chemistry contained in these meteorites indicates that the martian sulphur cycle includes movement of sulphur from the atmosphere, perhaps introduced by volcanic eruptions, back into surface materials (the authors' scheme is shown in their Fig. 3 on page 52). It is not clear how atmospheric sulphur was incorporated into solid rocks, but an apparent correlation with oxygen isotopes suggests hydrothermal activity as one possibility.



Figure 1 The surface of Mars, which is rich in sulphur. From isotope studies of meteorites, Farquhar and colleagues¹ conclude that the sulphur came from the martian atmosphere after originating, perhaps, from volcanic eruptions.

This work has another implication: chemical reactions in the atmosphere provide a way of bringing about large changes in the ³⁴S/³²S ratio without involving biological processes. Many scientists expect that the most likely evidence for past or present life on Mars might be found in the large isotopic fractionations produced by organisms. But the new results show that the ³⁴S/³²S biomarker must be used with caution on any samples returned from Mars. Moreover, fractionation of ³³S/³²S and ³⁶S/³²S should also be taken into account to ensure that the results of atmospheric processes are not confused with biological effects. ■

Harry Y. McSween Jr is in the Department of Geological Sciences, University of Tennessee, Knoxville, Tennessee 37996, USA.
e-mail: mcsween@utk.edu

1. Farquhar, J., Savarino, J., Jackson, T. L. & Thieme, M. H. *Nature* **404**, 50–52 (2000).
2. Rieder, R. *et al. Science* **278**, 1771–1774 (1997).
3. McSween, H. Y. *Meteoritics* **29**, 757–779 (1994).
4. Greenwood, J. P. *et al. Geochim. Cosmochim. Acta* (in the press).
5. Greenwood, J. P. *et al. Geochim. Cosmochim. Acta* **61**, 4449–4453 (1997).
6. Nielsen, H. in *Lectures in Isotope Geology* (eds Jäger, E. & Hunziker, J. C.) 283–312 (Springer, Berlin, 1979).
7. Thieme, M. H. *Science* **283**, 341–345 (1999).

Daedalus

Brief lives

'A dog is for life — not just for Christmas!' This slogan is intended to warn potential purchasers of the long-term consequences of pet ownership: not merely dogs, and not merely at Christmas. Yet thousands of households buy pets to amuse the children, and are stuck with them when the children are no longer amused. Now Daedalus has the answer. He is inventing the short-life pet.

In principle, no mammal can live for ever. At every cell division, one unit is lost from the telomere sequence on each of its chromosomes. When the entire telomere sequence has gone, the cell can no longer divide. Its vital reserve is exhausted. So Daedalus plans to shorten the telomere sequence in pet animals. His biologists are taking fertilized dog, cat and hamster ova, and cultivating them *in vitro* by standard methods.

Every animal, of course, arises from a single fertilized ovum. It divides into two cells, then into four, and ultimately into a complete fetus. If at the two-cell stage the cells are separated, they go on to develop into identical twins. The DREADCO team members are encouraging this process. Each time an ovum divides into two, they separate the two cells. When these reach the two-cell stage, these are separated again, and so on. At each cell-division, one telomere unit is lost from the cells' chromosome termination. After many such divisions, Daedalus will have millions of identical telomere-depleted ova of the chosen pet species. They will be implanted into surrogate mothers of the same species, and brought to term.

The resulting identical youngsters will command vast prices in the pet market. Children will adore their fluffy appeal. Canny parents will appreciate their brevity. Almost as soon as their appeal has waned, their telomere sequences will be exhausted, and they will drop dead.

Daedalus is not sure what they will die of. It won't be senility; brain cells hardly divide at all. Gut cells divide very fast, so it may be indigestion. But one standard feature of old age, in Man as in animals, is the increasing embrittlement and fragility of connective tissue. So their fibroblast telomeres may give out first. Like other toys, Daedalus's short-life pets may just fall to pieces.

David Jones

The Further Inventions of Daedalus (Oxford University Press), 148 past Daedalus columns expanded and illustrated, is now on sale. Special *Nature* offer: m.curtis@nature.com

Errors in judging 'offside' in football

Optical trickery can undermine the assistant referee's view of this ruling.

In football (soccer), a player is 'offside' if he or she is closer to the goal than the last defender (excluding the goalkeeper) when the ball is passed to them. We investigated why assistant referees, who have the responsibility of judging offside, regularly make mistakes. We show that this is probably due to the angle of viewing by the assistant referee, who is frequently positioned beyond the last defender — a viewpoint from which errors are optically inevitable.

In a field experiment, three professional assistant referees (ARs, also known as linesmen) judged 200 potential offside situations played by two elite youth football teams (Fig. 1a). The ARs made 40 errors. One explanation for these errors is that the AR cannot see passer and receiver simultaneously: this causes the AR to shift his gaze from passer to receiver and so make judgements a split-second after the moment of passing — long enough for the receiver to have gone past the last defender and to appear offside'. We found, however, that this is an unlikely explanation for these errors, because an AR equipped with a head-mounted camera showed no shift of gaze from passer to receiver.

In 179 situations, the assistant referee was positioned beyond the last defender (mean, 1.18 m; s.d. = 0.94). In Fig. 1b, the 'outside' attacker is not offside. However, when attacker and defender are projected onto the AR's retina, the image of the attacker is just to the right of that of the defender. This means that the attacker is perceived as being in front of the defender, prompting the AR to wrongly raise his flag to call offside (flag error, FE). By contrast, in Fig. 1c the outside attacker is offside. But the AR will perceive attacker and defender as being in line, and so keep his flag down (no-flag error, NFE).

If these ideas are correct, then, when the attacker goes outside the defender (Fig. 1b), more FEs than NFEs should occur

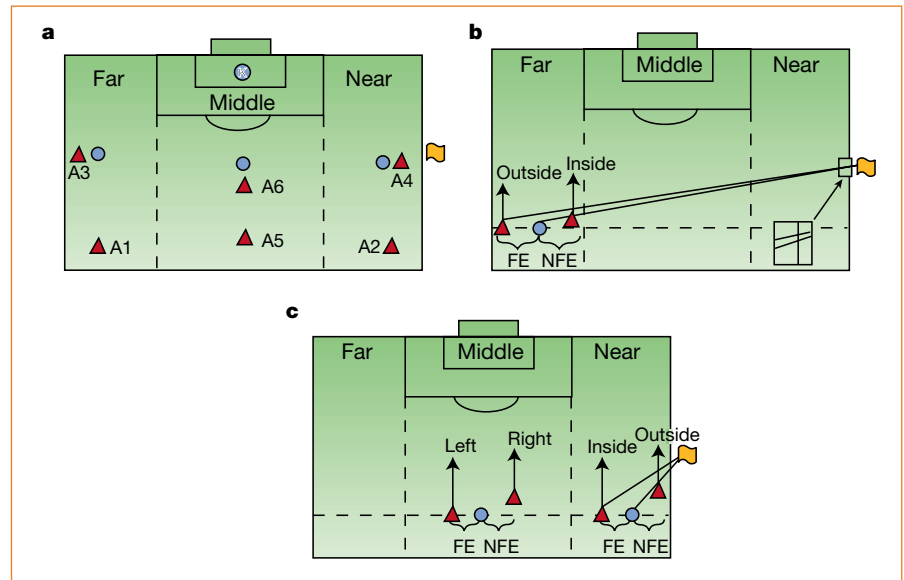


Figure 1 Offside situations. **a**, Experimental set-up of players in simulated offside situations. Attackers are triangles, circles are defenders (K is the goalkeeper, also known as the netminder), and the flag at the sideline depicts the position of the assistant referee (AR). Situations were videotaped with two cameras from an apartment block next to the pitch, so that we could determine whether the AR had judged the situation correctly. One AR wore a lightweight, head-mounted video camera to record his head movements relative to the scene. Experimental details are available from the authors. **b,c**, How the AR would perceive the relative positions of attacker and defender in far situations (**b**), and in situations near the AR and in the middle of the field (**c**). FE (flag error) and NFE (no-flag error) show the kinds of error expected in each situation.

when the players are on the far side of the pitch from the AR, whereas the converse would be expected to occur when they are close to the AR (Fig. 1). In contrast, when the attacker goes inside, more NFEs than FEs should occur far from the AR, and more FEs than NFEs should occur near the AR. This also holds for judging offside in the middle zone (Fig. 1c): when the attacker goes right, NFEs are expected, and when left, FEs.

Data from our experiments (first row of Table 1), and from 200 videotaped football matches from five national competitions (1996–98 seasons) and the 1998 World Cup (Table 1) confirmed these expectations. In situations far from the AR, more FEs than

NFEs were made when the attacker went outside the defender. In situations near the AR, more NFEs were made. If the attacker went past the defender on the inside, the opposite occurred. In the middle zone, there were 48 NFEs and 18 FEs when the attacker went right, and 61 FEs and 18 NFEs when they went left ($\chi^2 = 36.17$, $P < 0.0001$).

In conclusion, errors made by ARs in judging offside may often be the result of the relative optical projections of the players on the AR's retina. This means that, regardless of the quality of the AR, judgement errors are inevitable owing to the apparent limitations of our perceptual system. In our results, 9.3% of the AR's calls of offside were FEs. Given the high stakes in modern football, this incidence of (inevitable) errors suggests that alternative ways of judging offside should be developed, such as off-line analysis of video images taken from an adequate observation point.

Raoul R. D. Oudejans, Raymond Verheijen, Frank C. Bakker, Jeroen C. Gerrits, Marten Steinbrückner, Peter J. Beek

Institute for Fundamental and Clinical Human Movement Sciences, Faculty of Human Movement Sciences, Vrije Universiteit, Van der Boechorststraat 9, 1081 BT Amsterdam, The Netherlands
e-mail: r_r_d_oudejans@fbw.vu.nl

1. Sanabria, J. et al. *Lancet* 351, 268 (1998).

Table 1 Frequencies of flag errors and no-flag errors

	League	Games	Far from AR		Near AR		χ^2	P
			FE	NFE	FE	NFE		
Attacker goes	Experimental		23	5	3	9	12.06	<0.001
'outside'	Spain	50	43	7	6	19	28.29	<0.0001
	The Netherlands	50	40	6	4	21	34.60	<0.0001
	Italy	25	15	4	1	9	12.59	<0.0005
	England	25	16	4	1	10	14.41	<0.0005
	Germany	25	19	3	2	10	15.97	<0.0001
	World Cup '98	25	15	2	2	6	10.00	<0.005
Total			171	31	19	84	127.30	<0.0001
Total 'inside'		200	21	43	34	16	13.92	<0.0005

The inter-observer agreement for the analysed matches, computed on the basis of a selection of 12 matches watched by a second observer, was 93% (142 out of 152) for all potential offside situations and 90% (27 out of 30) for errors made by the ARs.

Bioinformatics

Finding genes in *Plasmodium falciparum*

The completion of the sequencing of chromosome 3 of the malarial parasite *Plasmodium falciparum*¹ is a major step forward in our understanding of the *Plasmodium* genome. We have analysed this chromosome using GlimmerM, a freely available gene-finder developed specifically for the *P. falciparum* species². GlimmerM was highly effective in finding nearly all the genes that were reported on *P. falciparum* chromosome 2 (ref. 2), and a newly re-trained version was even more effective on chromosome 3, confirming virtually all reported genes and finding several additional ones.

Using GlimmerM, we found 25 possible new genes on chromosome 3 in regions currently annotated as non-coding¹ (Table 1). Although some of these are relatively short and may not be genuine coding sequences, six of them are longer than 400 base pairs (bp) and one is 2,187 bp. This last gene represents a 729-amino-acid protein that is encoded by one very large exon and one shorter exon. Open reading frames of this length in a chromosome with 80% A+T content are virtually certain to represent real genes. Three of the table entries, G802, G803 and G740, have detectable homology to *var* gene fragments on chro-

somosome 2; G802 and G803 are close enough that they might represent two portions of the same gene.

Of the 215 protein-coding regions reported by Bowman *et al.*¹, GlimmerM automatically finds 214 of them. (An earlier version of GlimmerM was provided to the annotators of chromosome 3.) The only gene missed is a short hypothetical protein (PFC0360w, 114 amino acids) with no homology to any known gene¹. Finally, given that chromosome 3 is 12 per cent larger than chromosome 2, an extrapolation based on gene density would predict 234 genes on chromosome 3, consistent with our finding that additional genes could be present.

GlimmerM is very accurate at identifying splice sites, having been trained on a carefully curated set of experimentally confirmed introns from the *P. falciparum* genome². GlimmerM's predictions suggest different splice sites for 49 of the 215 genes annotated on chromosome 3; all of these are hypothetical proteins. As with the annotation of hypothetical proteins for chromosome 2, substantial additional laboratory studies are needed in order to determine with confidence the exon structure of these genes.

For chromosome 2, we used the polymerase chain reaction with reverse transcription for 13 hypothetical genes predicted by GlimmerM: all 13 predictions were confirmed^{2,3}. Of course, as emphasized

previously^{2,3}, GlimmerM is only one step, albeit an important one, in a process that should involve many other computational methods as well as careful human curation of the results produced by those methods.

To achieve the highest quality in analysing genome sequences, peer-reviewed bioinformatics methods should be used when they are available. We consider it an oversight that the chromosome 3 annotation effort neglected to consider the predictions of GlimmerM, especially as a substantial part of the chromosome 3 analysis involves comparing the two chromosomes (see, for example, Table 2 of Bowman *et al.*¹).

Mihaela Pertea, Steven L. Salzberg, Malcolm J. Gardner

The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, Maryland 20850, USA
e-mail: salzberg@tigr.org

1. Bowman, S. *et al.* *Nature* **400**, 532–538 (1999).
2. Salzberg, S. L. *et al.* *Genomics* **59**, 24–31 (1999).
3. Gardner, M. J. *et al.* *Science* **282**, 1126–1132 (1998).

Lawson *et al.* reply — Pertea *et al.* report their interpretation of the published *Plasmodium falciparum* chromosome 3 sequence¹ using The Institute for Genomic Research's gene-prediction algorithm GlimmerM². We believe, however, that their analysis is flawed, and highlights the problem that overreliance on a single algorithm can lead to mistakes in annotation.

The accurate *ab initio* prediction of a eukaryotic coding sequence is extremely difficult because of the presence of intronic DNA. Gene predictions are distillates of the results from several algorithms that differentiate coding from non-coding segments. The role of the annotator is to identify putative genes and achieve the best prediction using the tools available. Ultimately, predictions can only be validated by laboratory studies.

GlimmerM finds 25 new gene models, which Pertea *et al.* suggest could be an oversight in the original analysis. Predictions G802, G803 and G740 map to regions of chromosome 3 already assigned as *varC* pseudogenes, as the region of similarity is heavily frameshifted¹. We do not believe that the prediction of exons from within these regions correctly interprets the data. Until laboratory investigation can show that these regions are coding, the parsimonious (and correct) annotation is that these regions are pseudogenes.

Many gene-prediction algorithms are designed to have a minimum gene size (quite often 100 amino acids). These limits are implemented not because of any biological rationale, but because the inclusion of such open reading frames presents major difficulties in the interpretation of the data as most are likely to be non-coding. There is

Table 1 Possible new genes on chromosome 3

ORF	No. of exons	Length		Flanking genes		Strand
		BP	AA			
G802	1	450	150	PFC0010c	PFC0025c	–
G803	1	468	156	PFC0010c	PFC0025c	–
G815	1	210	70	PFC0030c	PFC0035w	+
G036	2	192	64	PFC0125w	PFC0130c	–
G070	1	220	73	PFC0175w	PFC0180c	–
G121	1	531	177	PFC0260w	PFC0265c	–
G124	1	411	137	PFC0260w	PFC0265c	–
G141	1	255	85	PFC0280c	PFC0285c	+
G144	4	348	116	PFC0280c	PFC0285c	+
G233	1	240	80	PFC0415c	PFC0420w	+
G256	1	459	153	PFC0440c	PFC0445w	–
G273	1	282	94	PFC0465c	PFC0470w	–
G282	2	249	83	PFC0480c	PFC0485w	–
G290	1	243	81	PFC0485w	PFC0490w	–
G294	1	285	95	PFC0490w	PFC0495w	+
G313	2	285	95	PFC0505c	PFC0510w	+
G356	2	293	131	PFC0555c	PFC0560c	–
G408	2	213	71	PFC0580c	PFC0590c	+
G410	1	231	77	PFC0580c	PFC0590c	+
G529	5	339	113	PFC0810c	PFC0815c	–
G612	3	249	83	PFC0910w	PFC0915w	–
G614	2	279	93	PFC0910w	PFC0915w	+
G702	2	2187	729	PFC1010w	PFC1015c	–
G734	1	237	79	PFC1060c	PFC1065w	+
G740	1	312	104	PFC1065w	PFC1070c	–

Analysis of chromosome 3 of *P. falciparum* by GlimmerM¹ generated 25 new gene models not included in current annotation. Columns show the number of exons in the predicted gene model, the length in base pairs (BP) and amino acids (AA), the nearest flanking genes on either side of the predicted new gene and the strand (+, forward strand; –, reverse complement). A version of this table, linked to the DNA and protein sequences for each gene, is available at <http://www.tigr.org/tldb/edb/pf3table.html>. ORF, open reading frame.

no indication that the 15 novel small predictions are coding, and the hexamer algorithm suggests that 12 of the 15 are non-coding. The remaining seven GlimmerM predictions form part of gene models that we consider to be borderline and are validating experimentally before inclusion in the chromosome annotation.

Pertea *et al.* suggest that peer-reviewed bioinformatics methods are essential. Certainly, if the diverse outputs of multiple gene-prediction algorithms can be included, then the annotation should be good, but that does not make the gene prediction correct — it remains a prediction and not a confirmed gene model.

Overreliance on the output of a predictive tool can lead to erroneous prediction and bad annotation. The well-established tools used in the analysis of chromosome 3, Genefinder, hexamer and ACEDB (P. Green and L. Hillier, unpublished software; R. Durbin, unpublished software) have a proven track record in eukaryotic gene prediction³.

Our analysis of chromosome 2 using Genefinder/hexamer indicates that 40 modifications need to be made to the original gene predictions (20 per cent of gene models). Although these new predictions require confirmation, they highlight the inaccuracy of first-pass annotation from uncharacterized genomic DNA. We encourage re-analysis of genomic data to increase sequence accuracy: Genefinder, hexamer and GlimmerM all have roles to play in the (re)annotation of *P. falciparum* chromosomes and in improving the interpretation of sequence data.

Dan Lawson, Sharen Bowman, Bart Barrell

Pathogen Sequencing Unit, The Sanger Centre,
Wellcome Trust Genome Campus,
Hinxton CB10 1SA, UK
e-mail: dl1@sanger.ac.uk

1. Bowman, S. *et al.* *Nature* **400**, 532–538 (1999).
2. Salzberg, S. L. *et al.* *Genomics* **59**, 24–31 (1999).
3. The *C. elegans* Sequencing Consortium *Science* **282**, 2012–2018 (1998).

Oceanography

Fish do not avoid survey vessels

The precarious condition of the world's fisheries is making ever-greater demands of the scientific assessment of fish stocks. Traditional assessments that rely on commercial catch statistics can have major shortcomings¹ (as shown, for example, by the collapse of Canada's northern cod stock²), increasing the need for more fishery-independent data. Acoustic surveys can provide such information³, but ocean-going research vessels have high operating

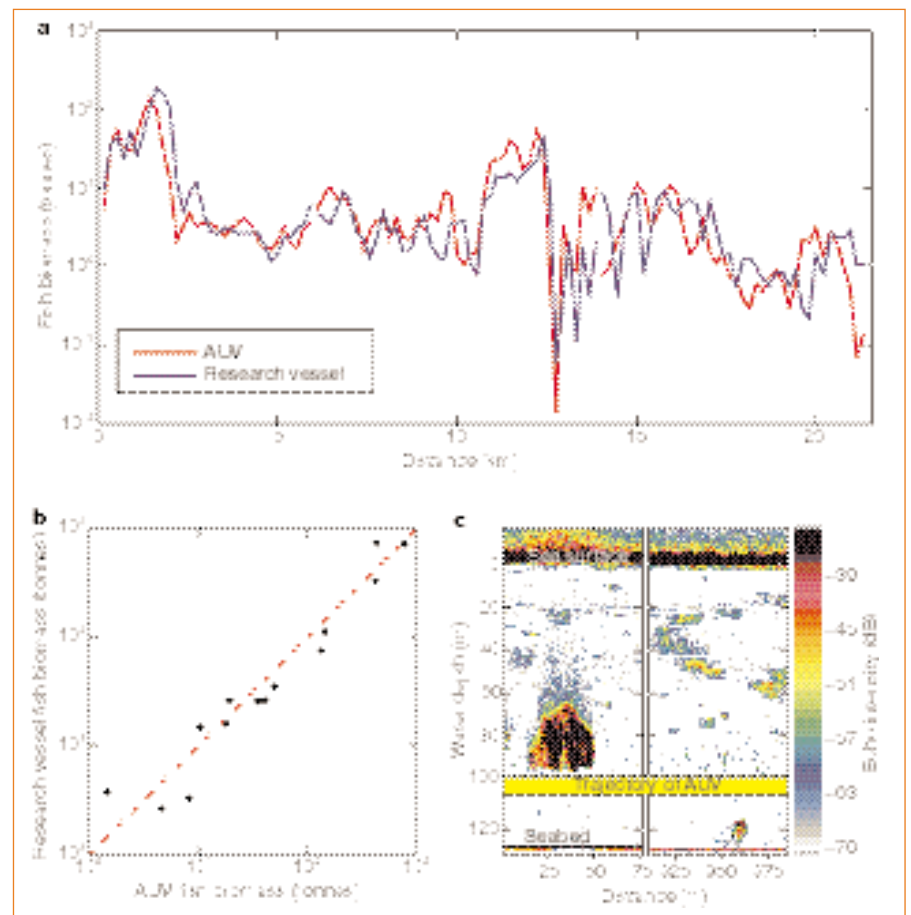


Figure 1 Comparison of acoustic data collected by an autonomous underwater vehicle (AUV) and a research vessel. **a**, Fish biomass at integrated intervals of 186 m along three transects on 21 July 1999; fish were identified as herring (*Clupea harengus*) by trawling. **b**, Strata-averaged biomass estimates from the AUV and research vessel were significantly correlated ($r=0.935$, $P<0.001$), lying around the (dotted) one-to-one line (Wilcoxon signed-rank test, $P>0.05$). **c**, Composite echogram collected by the AUV at 105 m (equipped with upward- and downward-looking transducers) as it passed less than 7 m beneath a large midwater herring school (left) and only 10 m above another close to the sea bed (right).

costs, and there is also widespread concern that fish avoid these vessels because of the noise they make, thereby biasing abundance estimates⁴. Here we present new data gathered by an autonomous underwater vehicle (AUV) showing that vessel avoidance is not a significant source of bias. Our investigation also heralds the arrival of AUVs as effective survey platforms.

During acoustic surveys, sound pulses are transmitted vertically downwards into the water at regular intervals (typically 1 s) from a survey vessel travelling along defined transects. Fish density is calculated by integrating the intensities of the returning echo⁵, and is then interpolated to give an estimate of the abundance in the survey area³.

We deployed the AUV *Autosub-1* (ref. 6) 200–800 m ahead of the research vessel *Scotia* on eight transects in water 60–180 m deep during an acoustic survey of herring in the North Sea. Herring have the most sensitive hearing of the commercially exploited species examined so far⁷ and are more likely than any to react to vessel noise. *Autosub-1* is unmanned and follows pre-programmed

mission trajectories. In comparison to the 68-m *Scotia*, it is small (torpedo shaped, 7 × 1 m) and extremely quiet (being propelled by an electric motor).

Avoidance of *Autosub-1* by herring is minimal: passing unprecedentedly close to a school (Fig. 1c), the vehicle caused only the localized school compression that typically occurs on close approach of predators⁸. *Autosub-1* was equipped with the same type of 38-kHz scientific echosounder as *Scotia*, and gathered equivalent acoustic data before the research vessel arrived. If fish avoided the *Scotia*, then it should have detected fewer fish than *Autosub-1*.

At the integrated resolution, there are small-scale temporal and spatial differences between the AUV and research-vessel data (Fig. 1a), but the underlying similarity in magnitude and trend is clear. For statistical comparison, data from each source were aggregated into independent strata of equivalent geographical area.

The amount of fish detected by the research vessel was not significantly different from that detected by the AUV (Fig. 1b). *Scotia* is very quiet, having been built

to guidelines intended to limit noise emission⁷. Our data show that for such vessels, avoidance is not a source of bias.

We also deployed *Autosub-1* on eight fully autonomous missions. This was the first time that any AUV had operated successfully beyond the control range of a support facility. These missions provided over 76 h (420 km) of additional survey data which, also for the first time, span the whole water column (Fig. 1c).

AUV technology is now sufficiently robust for effective fish-stock monitoring⁹, and could advance fisheries surveys by allowing acoustic detection closer to the target species. This could improve the assessment of groundfish, such as cod, and deep-water fish, and may also facilitate more extensive high-frequency zooplankton studies¹⁰. Reductions in battery costs and improvements in the acoustic identification of species using multifrequency¹¹ and broadband¹² techniques may eventually enable AUVs to replace research vessels as acoustic sampling platforms.

Autosub-1's unique capabilities are being used in a variety of other marine-science programmes¹³, from studies of ocean turbulence to marine geochemistry. We plan to exploit the vehicle's ability further to operate in otherwise impenetrable environments later this year, when we will deploy it under Southern Ocean sea-ice to measure ice thickness, and the abundance and distribution of Antarctic krill.

P. G. Fernandes*, A. S. Brierley†, E. J. Simmonds*, N. W. Millard‡, S. D. McPhail‡, F. Armstrong*, P. Stevenson‡, M. Squires‡

*FRS Marine Laboratory Aberdeen, PO Box 101, Victoria Road, Aberdeen AB11 9DB, UK

†British Antarctic Survey, High Cross, Madingley Road, Cambridge CB3 0ET, UK

‡Southampton Oceanography Centre, Empress Dock, Southampton SO14 3ZH, UK
e-mail: fernandespg@marlab.ac.uk

- Walters, C. J. *Western Fisheries* 41–44 (January, 1996).
- Rose, G. A. & Kulka, D. W. *Can. J. Fish. Aquat. Sci.* **56** (suppl. 1), 118–127 (1999).
- MacLennan, D. N. & Simmonds, E. J. *Fisheries Acoustics* (Chapman & Hall, London, 1992).
- Magurran, A. E. in *Dynamics of Pelagic Fish Distribution and Behaviour: Effects on Fisheries and Stock Assessment* (eds Freon, P. & Misund, O. A.) ix–x (Fishing News Books, Oxford, 1999).
- Foot, K. G. *J. Acoust. Soc. Am.* **73**, 1932–1940 (1983).
- Millard, N. W. *et al. J. Soc. Underwater Tech.* **23**, 7–17 (1998).
- Mitson, R. B. *ICES Co-op. Res. Rep.* **209**, 1–61 (1995).
- Freon, P., Gerlotto, F. & Soria, M. *Fish. Res.* **15**, 45–66 (1992).
- Fernandes, P. G. & Brierley, A. S. *ICES CM 1999/M:01*, 1–16 (1999).
- Holliday, D. V. & Pieper, R. E. *ICES J. Mar. Sci.* **52**, 279–296 (1995).
- Brierley, A. S., Ward, P., Watkins, J. L. & Goss, C. *Deep-Sea Res.* **45**, 1155–1173 (1998).
- Simmonds, E. J., Armstrong, F. & Copland, P. J. *ICES J. Mar. Sci.* **53**, 189–195 (1996).
- Millard, N. W. *et al. in Proc. Int. Conf. Remote Techniques for Hazardous Environments* (ed. Barnes, S.) 19–20 (Telford, London, 1999).

Evolution

Migration and speciation

Although migration is a common behaviour, the effects of this annual two-way event on the speciation process are poorly understood, even though birds, which are commonly migratory, played a critical role in the development of speciation theory^{1,2}. Here I propose that new developments^{3,4} in the theory of sympatric speciation — a process whereby new species can arise through population differentiation without spatial isolation — may help to explain the bursts of speciation observed in some seasonal migrant lineages.

Seasonal migrants, particularly those travelling long distances, have few recognized limits to their ability to disperse into new environments^{5,6}. These colonizations can foster the production of new species flocks — examples include nearctic *Catharus* thrushes (Fig. 1), some palaeartic *Phylloscopus* warblers, and the irruptive migrant specialists red crossbills (*Loxia curvirostra*)^{7–9}. I suggest that migration itself can be viewed as a key innovation that occasionally enables lineages to radiate in new environments.

Migration is a complex mode of dispersal, promoting the colonization of new areas, but also their regular re-colonization and gene flow. Spatial segregation — the linchpin of most speciation theory — becomes less and less likely with increasing migratory tendencies. Achieving true geographic isolation from other populations, thereby allowing differentiation to occur in the absence of gene flow, seems particularly unlikely among long-distance migrants, whose movements regularly encompass entire continents and oceans.

But here we have a conundrum: while migration opens the door to differentiation in new ecological and geographic space, it apparently slams it shut again through denial of geographic isolation and the promotion of gene flow.

Until now, migration was considered to counter differentiation¹⁰. Scenarios proposed to explain migrant speciation have had to invoke geographic isolation and, by implication, mechanisms such as lower historic levels of migration and greater levels of natal philopatry — neither of which fits the evidence^{11–13}. Although it is true that the origins and losses of migration have occurred independently in many lineages, it is unrealistic to suggest that the associated complex life-history characteristics were somehow held in temporary abeyance across entire lineages or clades.

Recent developments in speciation theory^{3,4} offer a theoretical framework to escape such ill-fitting scenarios, and species flocks of migrants could provide a testing ground for these theories. Phenotypic evidence in



Figure 1 Hermit thrush (*Catharus guttatus*; top) and Swainson's thrush (*C. ustulatus*; bottom), two nearctic-neotropical migrants from a highly migratory clade. These two species and three close relatives comprise a group of neotropical origin. They have trans-continental breeding distributions in temperate- and high-latitude nearctic forests, and wintering distributions centred on neotropical North and South America.

birds suggests that sexual selection may operate only as a distant second to resource competition and perhaps reinforcement (adaptations to prevent hybridization) in driving speciation events among many migratory animals.

Kevin Winker

University of Alaska Museum, 907 Yukon Drive, Fairbanks, Alaska 99775-6960, USA
e-mail: fjksw@uaf.edu

- Darwin, C. *On the Origin of Species* (John Murray, London, 1859).
- Mayr, E. *Animal Species and Evolution* (Belknap, Cambridge, MA, 1963).
- Kondrashov, A. S. & Kondrashov, F. A. *Nature* **400**, 351–354 (1999).
- Dieckmann, U. & Doebeli, M. *Nature* **400**, 354–357 (1999).
- Grinnell, J. *Auk* **39**, 373–380 (1922).
- Böhning-Gaese, K., González-Guzmán, L. I. & Brown, J. H. *Evol. Ecol.* **12**, 767–783 (1998).
- Helbig, A. *Ibis* **138**, 650–666 (1996).
- Richman, A. D. *Evolution* **50**, 767–783 (1996).
- Groth, J. G. *Univ. Calif. Publ. Zool.* **127**, 1–143 (1993).
- Montgomery, T. H. *Jr Am. Nat.* **30**, 458–464 (1896).
- Dilger, W. C. *Auk* **73**, 313–353 (1956).
- Mengel, R. M. *Living Bird* **3**, 9–43 (1964).
- Cox, G. W. *Am. Nat.* **126**, 451–474 (1985).

Correction

Arsenic poisoning in the Ganges delta

T. R. Chowdhury *et al.*; reply from J.M. McArthur
Nature **401**, 545–547 (1999)

The form of citation of one of the references given in this exchange (ref. 10 of Chowdhury *et al.*, ref. 6 of McArthur) was misleading and should have been written as "British Geological Survey/Mott MacDonald Ltd *Groundwater Studies for Arsenic Contamination in Bangladesh* (1999)." This is because it is the final report of Phase 1: *Rapid Investigation Phase*, not of the overall project, for which the final report will appear in 2000.

The importance of repairing stalled replication forks

Michael M. Cox^{*}, Myron F. Goodman[†], Kenneth N. Kreuzer[‡], David J. Sherratt[§], Steven J. Sandler^{||} & Kenneth J. Mariani[¶]

^{*} Department of Biochemistry, University of Wisconsin-Madison, Madison, Wisconsin 53706-1544, USA

[†] Departments of Biological Sciences and Chemistry, University of Southern California, Los Angeles, California 90089-1340, USA

[‡] Department of Microbiology, Duke University Medical Center, Durham, North Carolina 27710, USA

[§] Division of Molecular Genetics, Department of Biochemistry, University of Oxford, Oxford OX1 3QU, UK

^{||} Department of Microbiology, University of Massachusetts, Amherst, Massachusetts 01003, USA

[¶] Molecular Biology Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA

The bacterial SOS response to unusual levels of DNA damage has been recognized and studied for several decades. Pathways for re-establishing inactivated replication forks under normal growth conditions have received far less attention. In bacteria growing aerobically in the absence of SOS-inducing conditions, many replication forks encounter DNA damage, leading to inactivation. The pathways for fork reactivation involve the homologous recombination systems, are nonmutagenic, and integrate almost every aspect of DNA metabolism. On a frequency-of-use basis, these pathways represent the main function of bacterial DNA recombination systems, as well as the main function of a number of other enzymatic systems that are associated with replication and site-specific recombination.

In bacterial cells, replication forks often encounter template DNA damage that can inactivate the fork. This problem is not restricted to situations in which cells are stressed by ultraviolet irradiation or other damaging treatments. Instead, replication forks are routinely inactivated under normal aerobic growth conditions, where SOS is not induced and some SOS functions are not present. Here we briefly review the pathways for the reactivation of these replication forks as framed in proposals from a number of laboratories based on more than a decade of work¹⁻²⁴. The main conclusions of this intensive research are that (1) most, if not all, of the replication forks initiating at the bacterial origin, *oriC*, encounter DNA damage under normal growth conditions; (2) many of these encounters inactivate the replication forks; (3) reactivation of the fork requires DNA recombination functions, a system for the origin-independent restart of replication and additional enzymes to reverse potentially detrimental side products of recombination; and (4) the pathways for reactivating the fork under normal growth conditions are nonmutagenic. In effect, the reactivation of replication forks represents a major housekeeping function in bacteria. Here we focus attention on the nonmutagenic pathways for replication fork reactivation and the enzymes involved in them, as well as highlighting the important contribution this process makes to cellular DNA metabolism during normal bacterial growth.

Figure 1 presents a few of the current ideas for replication fork demise and nonmutagenic reactivation, although we acknowledge at the outset that many of the molecular details of these pathways remain to be elucidated. The replication forks originating at *oriC* include the DNA polymerase III holoenzyme and a primosome consisting of DnaB and DnaG. An encounter with DNA damage results in an enzymatic train wreck that we refer to as either replication fork demise or inactivation. This is a working definition, underscoring that replication fork progression has been arrested. It is not yet known to what extent the replication machinery disassembles in these situations, and the molecular consequences may be as varied as the types of damage encountered.

The two main possibilities for fork demise that we consider are, first, if a strand break is encountered, a double-strand break will be generated. And, second, if an unrepaired DNA lesion is encountered, the replication fork halts and a DNA gap will be created. In broad strokes, the events required to reactivate the replication fork in both cases involve recombination (by at least two major path-

ways), replication restart, completion of elongation of nascent chains and resolution of any dimeric chromosomal products that result from recombination (Fig. 2). The elaborate molecular requirements for reactivating a replication fork imply a high degree of coordination between recombination and replication functions,

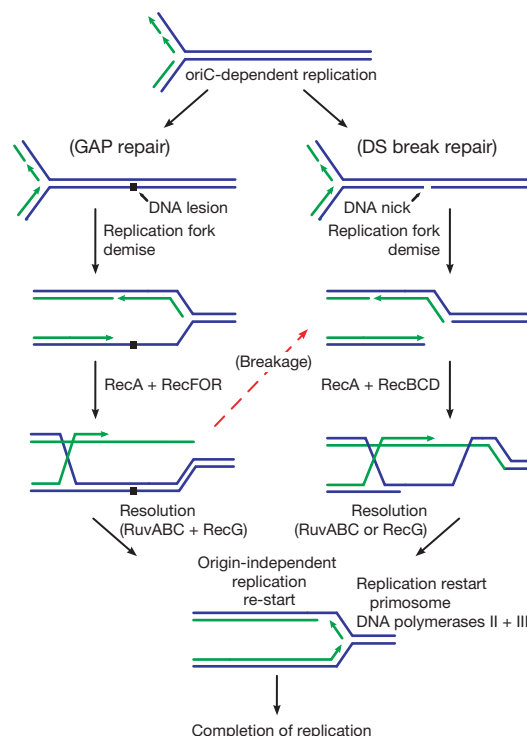


Figure 1 Some potential pathways for the nonmutagenic re-establishment of inactivated replication forks in bacteria. The pathways shown illustrate two of the important situations during normal cell growth that may result in replication fork demise, encounter with a DNA lesion or a DNA strand break. Reactivation involves the two main homologous genetic recombination pathways. The processes shown are broadly based on some published studies and discussions at recent national meetings; however, many of the details shown are speculative. The configurations of DNA strands shown in the intermediates are neither representative of all the proposals for fork reactivation nor intended to represent anyone's ideas of the most likely paths.

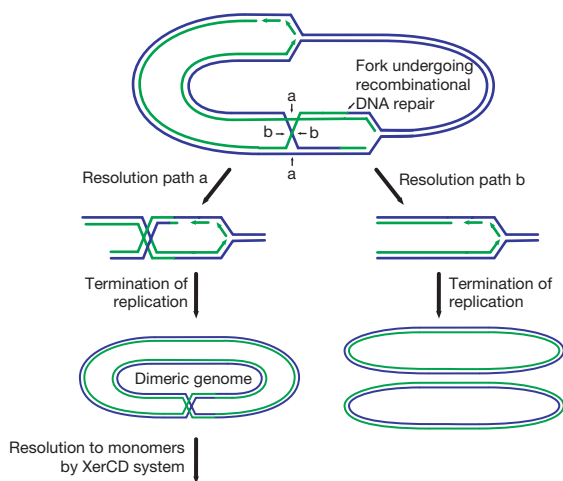


Figure 2 Creation and resolution of contiguous chromosomal dimers as a byproduct of the recombination required for re-establishment of inactivated replication forks. If Holliday junctions are created upstream from the replication fork during the repair process, their resolution can result in the creation of chromosomal dimers that must be resolved by the XerCD site-specific recombination system.

although the biochemical interactions between the replication and recombination systems remain mostly unexplored.

The entire process may involve more than two dozen proteins (Table 1), and only small parts have been reconstituted *in vitro*. Notably, the process includes important functions for several enzymatic systems whose functions have been, until recently, obscure. These include (1) the replication restart primosome, a complex of seven proteins discovered during investigation of the replication of bacteriophage ϕ X174 DNA *in vitro*^{25,26}, but later found to be unnecessary for *oriC*-dependent replication²²; (2) DNA polymerase II, discovered almost 30 years ago, but until recently without a significant identified function in bacterial DNA metabolism¹⁸; (3) the XerCD site-specific recombination system¹⁷; and (4) the RecF, O and R proteins, for which the associated recombination pathways have generally not been apparent unless other pathways are mutationally altered or removed²⁷.

The general concepts of bacterial replication fork demise and reactivation are not new, but have evolved during the work of multiple laboratories over a period of nearly 35 years. The first suggestion that a replication fork might be inactivated at the site of a strand break came in 1966 (ref. 1). On the basis of studies of phage lambda DNA replication, Shalka² later argued that pre-existing nicks in the parental template can lead to replication fork demise, and also that replication forks might be reactivated by a recombinational process. A link between recombination and DNA repair was evident in the phenotypes of the first recombination mutants^{28,29}. Some still-viable proposals for the recombinational repair of DNA gaps and double-strand breaks after the demise of a replication fork are decades old^{2,3}. An early model for the integration of recombination and replication was provided by the work of Mosig, Alberts and their colleagues on the bacteriophage T4 system^{4,5}. Indeed, the recombination-dependent and nonmutagenic initiation of replication is needed for repair of inactivated replication forks and has a major role in the life cycle of bacteriophages such as T4 (ref. 16). The first recognition that replication fork demise (and a need for reactivation) might be commonplace in bacteria under normal growth conditions arose from studies of the phenotypes of *priA* mutants in 1991 (refs 16, 30).

More recently, Kuzminov⁷ consolidated many disparate results regarding *Escherichia coli* replication and recombination into a clear model for reactivating bacterial replication forks that had been inactivated upon an encounter with a template strand break, and argued that the main role of *E. coli* recombination proteins is to reconstitute inactivated forks. Kuzminov³¹ also argued that stalled replication forks are subject to an active breakage process (followed by reactivation). Michel and colleagues¹⁴ provided experimental evidence for this directed breakage event, as well as insight to the frequency at which double-strand breaks occur under normal growth conditions. The work of Kogoma and colleagues also played a key role in the development of these ideas. Building on extensive studies of the recombination-dependent replication observed during the SOS response^{8,10}, Kogoma¹⁰ pointed out that these processes could provide a general pathway to reactivate replication forks. Studies elucidating the function of the PriA protein^{6,11,32}, other primosomal components^{9,33,34} and the XerCD site-specific recombination system^{15,17} have been critical in developing the idea that replication fork reactivation is a housekeeping function of bacterial cells that operates at high frequency under normal growth conditions. More complete accounts of these works

Table 1 Proteins that may participate in replication fork reactivation

Protein*	Main activity	Role in fork reaction
RecA	Strand pairing/exchange	Formation of joint molecules
RecBCD	Chi site-modulated nuclease	Initiation of repair at double-strand breaks
RecF	Binds single- and double-stranded DNA	Limits extension of RecA filament
RecO	Binds SSB-coated DNA	Facilitates RecA loading to SSB-coated DNA
RecR	Binds SSB-coated DNA	Facilitates RecA loading to SSB-coated DNA
RuvAB	Branch migration DNA helicase	Resolution of joint molecules
RecG	Branch migration DNA helicase	Modulate structure of stalled fork?
RuvC	Holliday junction endonuclease	Resolution of joint molecules
RusA	Holliday junction endonuclease	Resolution of joint molecules
PriA	3' → 5' DNA helicase, binds bent DNA	Initiates replication fork assembly
PriB	Facilitates complex formation between PriA and DnaT	Replication fork assembly
PriC	Primosome assembly	Replication fork assembly
DnaT	Primosome assembly	Replication fork assembly
DnaB	5' → 3' DNA helicase	Replication fork helicase
DnaC	Binds and loads DnaB to DNA	Replication fork assembly
DnaG	Primase	Okazaki fragment primase
Rep	3' → 5' DNA helicase	Processing of stalled fork?
DNA polymerase III holoenzyme	DNA polymerase	Replicative polymerase
DNA polymerase II	DNA polymerase	Replication-restart at template lesions
DNA polymerase I	DNA polymerase, 5' → 3' exonuclease	Gap sealing
DNA ligase	Ligase	Gap sealing
SSB	Single-stranded DNA-binding protein	Coats single-stranded DNA
XerCD	Site-specific resolvase	Resolves dimeric chromosomes

* This listing does not include proteins that are known or suspected to be involved in replication for which a biochemical function has not been elucidated.

and additional syntheses can be found in a number of recent reviews^{10,22–24,35}.

Replication fork reactivation is required often

Quantitative estimates for the frequency of replication fork demise and reactivation are gradually becoming available. These have been reviewed^{23,36} and will only be summarized here. Chromosomal dimers are found in bacterial cells under normal growth conditions. Although these dimers might sometimes arise for other reasons, the vast majority of them almost certainly result from the recombinational repair events required to reactivate replication forks^{15,37} (Fig. 2). Studies by Kuempel and co-workers^{15,37} on the frequency of chromosomal dimer resolution by the XerCD recombinase thus provides a useful lower limit for the frequency of replication fork reactivation. Under normal growth conditions, about 15% of bacterial chromosomes undergo recombination required for reactivation of a replication fork to generate a crossover leading to formation of a contiguous chromosomal dimer. If resolution of Holliday junction recombination creates crossovers 50% of the time, this measurement only detects half of the fork reactivation events. The real frequency of fork reactivation by these pathways might be different if the Holliday junction resolution is skewed to favour either crossovers or non-crossovers.

Further estimates come from studies of mutants affecting enzymes involved in fork reactivation. Elimination of RecBCD results in the appearance of unrepaired double-strand breaks, presumably arising from inactivation and cleavage of replication forks, in 15–20% of the cells under normal growth conditions^{13,14}. Many other studies suggest that fork reactivation is yet more frequent. Up to 50% of cultured *recA* cells are dead and a substantial number of chromosomes have been lost³⁸. Viability of cells lacking PriA function is even lower^{9,22}. Both *recA* and *priA* null mutants become inviable when paired with mutations in a number of other recombination functions.

As reviewed elsewhere^{10,22–24,35}, the most prominent phenotypic effects of *rec* and *pri* mutants are closely linked to DNA damage under conditions in which cells are actively replicating DNA. The requirements for the *rec* and *pri* functions in reactivating replication forks provide an internally consistent explanation for the phenotypes of cells lacking one or more of these functions. At present, there are no alternative hypotheses that can explain the demonstrated importance of these genes to cell viability under normal growth conditions *in vivo*.

Replication fork reactivation encompasses redundant pathways that adapt to whatever DNA structure is presented to the cell at the site of an inactivated replication fork, with some of these outlined in Fig. 1. Because of this redundancy, the effects of some single mutations inactivating a component of these pathways are deceptively modest. However, bacterial cells are inviable if all avenues for replication fork reactivation are removed (by mutation of required genes), and this in turn implies that most replication forks must be reactivated at some point during their journey from *oriC* to the terminus. Certain proteins such as RecA and PriA seem to be required for the main reactivation pathways. However, the lack of complete inviability seen for individual mutations suggests that not all pathways require RecA, PriA and/or RuvC. For example, preliminary results suggest that a PriA-independent pathway for replication fork reactivation might involve PriC and the Rep helicase, because the *priA priC* (S. J. S., unpublished data) and *priA rep*¹⁴ double mutants are both inviable. On the other hand, the *priC rep* double mutant is viable (S. J. S., unpublished data).

The evident prevalence of replication fork demise and reactivation may require a readjustment of the commonly cited rate of replication fork propagation of 1,000 nucleotides per second that has been based on the time required for completion of one round of chromosomal replication³⁹. Replication fork reactivation clearly takes time. Estimates extrapolated from the reconstitution of

small parts of these systems *in vitro*, as well as observations *in vivo*, range from 15–50 minutes^{40–42}. Thus, it would seem that the fork has to be able to move faster than the previously calculated estimate.

Relationship to other systems

The concept of replication fork demise and reactivation under normal growth conditions owes much to studies in other areas. Work on bacterial recombination and unusual modes of replication has largely focused either on the SOS response or on bacterial conjugation, where important phenomena are amplified.

Some of the effects of replication fork demise can be more readily studied when all replication is synchronously halted, as occurs when artificially elevated levels of DNA damage lead to the induction of the SOS system⁴³. However, the resulting data are complicated by the induction of new pathways for replication that appear to be unique to the SOS response. SOS includes a number of downstream processes, such as cell-cycle arrest induced by SulA and the specialized mutagenic repair brought about by DNA polymerases IV (DinB) and V (UmuD₂C)-mediated replication fork bypass^{44–46}. These can be considered extreme measures that evolved to maximize cell survival under conditions where many cells are destroyed. SulA, DinB and UmuD₂C are functions present at significant levels only when SOS is induced. UmuD is present in cells under normal growth conditions (mostly in the unactivated form), but UmuC is not detectable⁴⁷. The nonmutagenic pathways for replication fork reactivation almost certainly operate during SOS, particularly at early stages before the full induction of the mutagenic paths. However, the presence of SOS-specific avenues for replication can make it difficult for studies under SOS conditions to illuminate the nonmutagenic pathways featured under normal growth conditions.

Not all of the SOS-specific replication processes are inherently mutagenic. As demonstrated primarily by Kogoma and colleagues^{8,10}, a form of replication, called inducible stable DNA replication (iSDR), which requires neither ongoing protein synthesis nor a functional *oriC*, is also induced during the SOS response. iSDR requires recombination functions as well as the replication restart primosome. To categorize the manner in which replication was initiating during iSDR, Kogoma and colleagues¹⁰ coined the term RDR, for recombination-dependent replication, defining it as “homologous recombination function-dependent replication triggered by a duplex DNA end.” RDR can be observed in many contexts, including bacterial conjugation, the replication of bacteriophage T4, unusual modes of replication during SOS and other processes. iSDR itself is a specialized cellular function and initiates only at unique origins of replication, of which there are two major ones, *iriM1* and *oriM2*, and requires a hypothetical SOS-induced endonuclease to make the initiating double-strand break^{8,10}. iSDR also does not process inactivated replication forks at the site of their demise; rather, it effects rescue by reinitiating synthesis of the entire chromosome. However, the study of iSDR and its requirements has contributed to evidence for fork reactivation pathways that are also utilized under normal growth conditions.

The work on conjugation has been critical in defining recombination pathways and discovering recombination functions. In particular, the RecBCD pathway for recombination, representing one of the principal avenues for reactivation of replication forks (Fig. 1), was mainly defined in studies of conjugation-associated recombination. In normal bacterial populations, however, conjugation events are typically separated by tens of thousands of cell generations.

Both conjugation and SOS are special situations that illuminate, but do not accurately reflect, the normal condition in bacterial cells. The nonmutagenic pathways for reactivating replication forks under normal growth deserve increased experimental attention. If replication fork reactivation is required in virtually every cell in every cell generation, then it is straightforward to conclude that the

processes outlined in Fig. 1 represent the main function of all of the enzymes listed here and in Table 1, with the exception of DNA polymerase III. Thus, the nonmutagenic re-establishment of inactivated replication forks is an essential housekeeping function. It includes all pathways of recombination (with or without double-strand breaks), replication restart and resolution of any chromosomal dimers that result through XerCD-mediated site-specific recombination.

It is important to note that reactivation of the replication fork does not necessarily require the repair of the lesion that caused its initial demise. Strand breaks will be repaired as a consequence of the recombination processes (Fig. 1), but lesions may simply be left behind in regions of now double-stranded DNA. They can subsequently be repaired by the excision repair or other repair systems, or they may cause additional problems during the next replication cycle. These repair systems may also be integrated into the fork reactivation pathways, but the extent of such integration (if any) remains to be determined.

Enzymatic conundrums in DNA metabolism resolved

The nonmutagenic reactivation of replication forks provide a *raison d'être* for a number of enzymatic systems that were previously considered enigmatic. Studies of the initiation of bacterial replication at *oriC* left no evident role for PriA and several associated proteins in what had been defined as the ϕ X174-type primosome^{22,48}. Nevertheless, cells deficient in PriA function are only marginally viable, and possess only one-fiftieth and one-eightieth of the wild-type capacity for recombination and the repair of ultraviolet-damaged DNA, respectively^{30,49}. This fact can be readily explained by the role of PriA in replication fork reactivation³³. Because the role of the ϕ X174-type primosome in the cell has now been clarified, it has been proposed³⁵ that it be referred to as the "replication restart primosome" (Fig. 1).

Xer site-specific recombination functions to resolve chromosomal dimers, and the importance of this resolution is readily understood within the context of replication fork reactivation. Several studies suggest that the XerCD system is well integrated into DNA metabolism and the cell cycle^{50,51}.

Recently, the 'orphan' DNA polymerase II has also found an important function *in vivo*, with a pivotal role in pathways for nonmutagenic initiation of replication restart during SOS, and probably in normal cells as well¹⁸. Here it appears that the restart primosome is required to initiate Okazaki fragments in the pol II replication restart pathway (M. F. G., unpublished data) and that pol II is then replaced subsequently by a reassembled replication complex containing pol III (ref. 18).

The recombination functions of the RecF pathway, particularly RecF, O and R, are required for conjugational recombination only in the absence of both RecBCD (required for recombinational repair of double-strand breaks) and SbcBC functions²⁷ and hence have also seemed superfluous. Within the context of replication fork reactivation, however, there is ample evidence that the RecBCD and RecFOR pathways of recombination are of similar importance to the cell^{15,23,52}.

Analogues in other systems?

The problems faced by *E. coli* in completing replication of its genome are probably faced by all organisms. Do similar processes exist in other bacteria, viral systems and even eukaryotic cells?

The use of a single replication origin and terminus region in bacterial systems would seem to place a premium on reactivation pathways to deal with forks that don't make the traverse of the entire genome successfully. Thus, perhaps the safest prediction is that many, if not all, bacteria will have similar systems. Key components such as RecA, XerCD and PriA are highly conserved among bacterial species. Interestingly, however, the complexity of the multiple fork reactivation pathways may vary. Many bacteria lack obvious PriB

and PriC homologues. We have already noted the importance of recombination-dependent replication for the life cycles of bacteriophage such as T4, as well as for replication fork reactivation.

Multiple origins, the myriad layers of control on initiation and the existence of multiple checkpoint systems make prognostication of the existence of fork reactivation systems in eukaryotes more problematic. Nevertheless, several groups have recently presented evidence that portions of eukaryotic chromosomes can be replicated by a recombination-dependent mechanism that may reflect nonmutagenic replication fork reactivation. In *Saccharomyces cerevisiae*, an internal chromosome break can be repaired by a process that very probably involves the establishment of a new replication fork through recombination with an intact homologue⁵³. The new replication fork then traverses the length of the chromosome out to the telomere^{54,55}. In addition, the successful replication of chromosomal ends in the absence of functional telomerase is dependent on recombination proteins^{56,57}. A surprising and very elegant study⁵⁸ has recently shown that the normal structure of telomeres in higher eukaryotic cells involves a protein-mediated, folded-back D-loop, with the distal tip of the chromosome invading an internal repeat of the same sequence; this is precisely the structure that initiates replication in recombination-dependent replication pathways that depend on double-strand breaks. It has also been suggested that recombination-directed replication can account for nonreciprocal chromosome translocations^{59,60}. Another study has provided a first look at the effects of encounters with a DNA lesion by an SV40 replication fork⁶¹.

Questions

Research to date has only scratched the surface of the causes of replication fork demise and the pathways of subsequent reactivation. Although replication can clearly be halted by DNA damage, there is little information concerning the molecular events associated with the demise of a replication fork. What are the specific types of damage present as a function of growth conditions? Does either the type of DNA lesion or its location (that is, leading- or lagging-strand template) dictate the pathway of replication fork reactivation taken? What does an inactivated replication fork look like? Are all of the twenty or so proteins at the replication fork lost, requiring reassembly of an entirely new complex? Does the restart primosome stay together until chromosome replication is completed, or is it replaced by the DnaB–DnaG primosome used at *oriC*? What transpires at the recombination–replication interfaces? Reconstitution of even one of the pathways for fork reactivation promises to be a major enzymological challenge.

In eukaryotic cells, DNA damage can induce one of several cell-cycle checkpoints. Irradiated cells can, for example, either experience a lengthening of S phase or be delayed at the G2/M boundary until the damage is repaired. A recent proposal suggests that a similar checkpoint, mediated by the SOS-inducible *umuDC* gene products, exists in *E. coli*⁶². Fork reactivation under normal growth conditions in bacterial cells may require a checkpoint system as well (J. McCool and S. J. S., unpublished data).

Replication forks have a range limited by DNA damage. If any one of the events outlined in Fig. 1 fails to take place, the affected cell will either die or undergo an aberrant cell-division event. This potential catastrophe may have provided the selective pressure needed for the evolution of homologous recombination systems and other enzymatic components needed for fork reactivation, a critical step paving the way for the evolution of organisms with larger genomes. □

1. Hanawalt, P. C. The U.V. sensitivity of bacteria: its relation to the DNA replication cycle. *Photochem. Photobiol.* **5**, 1–12 (1966).
2. Skalka, A. in *Mechanisms in Recombination* (ed. Grell, R. F.) 421–432 (Plenum, New York, 1974).
3. West, S. C., Cassuto, E. & Howard-Flanders, P. Mechanism of *E. coli* RecA protein directed strand exchanges in post-replication repair of DNA. *Nature* **294**, 659–662 (1981).

4. Luder, A. & Mosig, G. Two alternative mechanisms for initiation of DNA replication forks in bacteriophage T4: priming by RNA polymerase and by recombination. *Proc. Natl Acad. Sci. USA* **79**, 1101–1105 (1982).
5. Formosa, T. & Alberts, B. M. DNA synthesis dependent on genetic recombination: characterization of a reaction catalyzed by purified bacteriophage T4 proteins. *Cell* **47**, 793–806 (1986).
6. Zavitz, K. H. & Mariani, K. J. Dissecting the functional role of PriA protein-catalysed primosome assembly in *Escherichia coli* DNA replication. *Mol. Microbiol.* **5**, 2869–2873 (1991).
7. Kuzminov, A. Collapse and repair of replication forks in *Escheria coli*. *Mol. Microbiol.* **16**, 373–384 (1995).
8. Kogoma, T. Recombination by replication. *Cell* **85**, 625–627 (1996).
9. Sandler, S. J., Samra, H. S. & Clark, A. J. Differential suppression of priA2:kan phenotypes in *Escherichia coli* K-12 by mutations in priA, lexA, and dnaC. *Genetics* **143**, 5–13 (1996).
10. Kogoma, T. Stable DNA replication: interplay between DNA replication, homologous recombination, and transcription. *Microbiol. Mol. Biol. Rev.* **61**, 212–238 (1997).
11. McGlynn, P., Al, D. A., Liu, J., Mariani, K. J. & Lloyd, R. G. The DNA replication protein PriA and the recombination protein RecG bind D-loops. *J. Mol. Biol.* **270**, 212–221 (1997).
12. Cox, M. M. Recombinational crossroads—eukaryotic enzymes and the limits of bacterial precedents. *Proc. Natl Acad. Sci. USA* **94**, 11764–11766 (1997).
13. Michel, B., Ehrlich, S. D. & Uezst, M. DNA double-strand breaks caused by replication arrest. *EMBO J.* **16**, 430–438 (1997).
14. Seigneur, M., Bidnenko, V., Ehrlich, S. D. & Michel, B. RuvAB acts at arrested replication forks. *Cell* **95**, 419–430 (1998).
15. Steiner, W. W. & Kuempel, P. L. Sister chromatid exchange frequencies in *Escherichia coli* analyzed by recombination at the dif resolvase site. *J. Bacteriol.* **180**, 6269–6275 (1998).
16. Mosig, G. Recombination and recombination-dependent DNA replication in bacteriophage T4. *Annu. Rev. Genet.* **32**, 379–413 (1998).
17. Neilson, L., Blakely, G. & Sherratt, D. J. Site-specific recombination at dif by *Haemophilus influenzae* XerC. *Mol. Microbiol.* **31**, 915–926 (1999).
18. Rangarajan, S., Woodgate, R. & Goodman, M. F. A phenotype for enigmatic DNA polymerase II: a pivotal role for pol II in replication restart in UV-irradiated *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **96**, 9224–9229 (1999).
19. Bidnenko, V. *et al.* sbcS sbcC null mutations allow RecF-mediated repair of arrested replication forks in rep recBC mutants. *Mol. Microbiol.* **33**, 846–857 (1999).
20. Saveson, C. J. & Lovett, S. T. Tandem repeat recombination induced by replication fork defects in *Escherichia coli* requires a novel factor, RadC. *Genetics* **152**, 5–13 (1999).
21. McGlynn, P. & Lloyd, R. G. RecG helicase activity at three- and four-strand DNA structures. *Nucleic Acids Res.* **27**, 3049–3056 (1999).
22. Mariani, K. J. PriA: at the crossroads of DNA replication and recombination. *Prog. Nucleic Acids Res. Mol. Biol.* **63**, 39–47 (1999).
23. Cox, M. M. Recombinational DNA repair in bacteria and the RecA protein. *Prog. Nucleic Acids Res. Mol. Biol.* **63**, 310–366 (1999).
24. Kuzminov, A. Recombinatorial repair in *Escherichia coli*. *Microbiol. Mol. Biol. Rev.* **63**, 751–813 (1999).
25. Arai, K. *et al.* Enzyme studies of phi X174 DNA replication. *Prog. Nucleic Acid Res. Mol. Biol.* **26**, 9–32 (1981).
26. Mariani, K. J. Enzymology of DNA in replication in prokaryotes. *CRC Crit. Rev. Biochem.* **17**, 153–215 (1984).
27. Clark, A. J. & Sandler, S. J. Homologous genetic recombination: the pieces begin to fall into place. *Crit. Rev. Microbiol.* **20**, 125–142 (1994).
28. Clark, A. J. & Margulies, A. D. Isolation and characterization of recombination-deficient mutants of *Escherichia coli* K12. *Proc. Natl Acad. Sci. USA* **53**, 451–459 (1965).
29. Howard-Flanders, P. & Theriot, L. Mutants of *Escherichia coli* K-12 defective in DNA repair and in genetic recombination. *Genetics* **53**, 1137–1150 (1966).
30. Nurse, P., Zavitz, K. H. & Mariani, K. J. Inactivation of the *Escherichia coli* priA DNA replication protein induces the SOS response. *J. Bacteriol.* **173**, 6686–6693 (1991).
31. Kuzminov, A. Instability of inhibited replication forks in *E. coli*. *Bioessays* **17**, 733–741 (1995).
32. Liu, J. & Mariani, K. J. PriA-directed assembly of a primosome on D loop DNA. *J. Biol. Chem.* **274**, 25033–25041 (1999).
33. Liu, J. I., Xu, L. W., Sandler, S. J. & Mariani, K. J. Replication fork assembly at recombination intermediates is required for bacterial growth. *Proc. Natl Acad. Sci. USA* **96**, 3552–3555 (1999).
34. Sandler, S. J. *et al.* dnaC mutations suppress defects in DNA replication and recombination associated functions in *priB* and *priC* double mutants in *E. coli* K-12. *Mol. Microbiol.* **34**, 91–101 (1999).
35. Sandler, S. J. & Mariani, K. J. Role of PriA in replication fork reactivation in *Escherichia coli*. *J. Bacteriol.* **182**, 9–13 (2000).
36. Cox, M. M. A broadening view of recombinatorial DNA repair in bacteria. *Genes to Cells* **3**, 65–78 (1998).
37. Steiner, W. W. & Kuempel, P. L. Cell division is required for resolution of dimer chromosomes at the dif locus of *Escherichia coli*. *Mol. Microbiol.* **27**, 257–268 (1998).
38. Skarstad, K. & Boye, E. Degradation of individual chromosomes in recA mutants of *Escherichia coli*. *J. Bacteriol.* **75**, 5505–5509 (1993).
39. Mariani, K. J. Prokaryotic DNA replication. *Annu. Rev. Biochem.* **61**, 673–719 (1992).
40. Anderson, D. G. & Kowalczykowski, S. C. The translocating RecBCD enzyme stimulates recombination by directing RecA protein onto ssDNA in a chi-regulated manner. *Cell* **90**, 77–86 (1997).
41. Courcelle, J., Carswell-Crumpton, C. & Hanawalt, P. *recF* and *recR* are required for the resumption of replication at DNA replication forks in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **94**, 3714–3719 (1997).
42. Kuzminov, A. & Stahl, F. W. Double-strand end repair via the RecBC pathway in *Escherichia coli* primes DNA replication. *Genes Dev.* **13**, 345–356 (1999).
43. Shinagawa, H. SOS response as an adaptive response to DNA damage in prokaryotes. *Exs* **77**, 221–235 (1996).
44. Tang, M. *et al.* Biochemical basis of SOS-induced mutagenesis in *Escherichia coli*: reconstitution of in vitro lesion bypass dependent on the UmuD₂C mutagenic complex and RecA protein. *Proc. Natl Acad. Sci. USA* **95**, 9755–9760 (1998).
45. Tang, M. J. *et al.* UmuD₂C is an error-prone DNA polymerase, *Escherichia coli* pol V. *Proc. Natl Acad. Sci. USA* **96**, 8919–8924 (1999).
46. Wagner, J. *et al.* The dinB gene encodes a novel *E. coli* DNA polymerase, DNA pol IV, involved in mutagenesis. *Mol. Cell.* **4**, 281–286 (1999).
47. Woodgate, R. & Ennis, D. G. Levels of chromosomally encoded Umu proteins and requirements for in vivo UmuD cleavage. *Mol. Gen. Genet.* **229**, 10–16 (1991).
48. Kaguni, J. M. & Kornberg, A. Replication initiated at the origin (*oriC*) of the *E. coli* chromosome reconstituted with purified enzymes. *Cell* **38**, 183–190 (1984).
49. Kogoma, T., Cadwell, G. W., Barnard, K. G. & Asai, T. The DNA replication priming protein, PriA, is required for homologous recombination and double-strand break repair. *J. Bacteriol.* **178**, 1258–1264 (1996).
50. Steiner, W., Liu, G., Donachie, W. D. & Kuempel, P. The cytoplasmic domain of FtsK protein is required for resolution of chromosome dimers. *Mol. Microbiol.* **31**, 579–583 (1999).
51. Recchia, G. D., Aroyo, M., Wolf, D., Blakely, G. & Sherratt, D. J. FtsK-dependent and -independent pathways of Xer site-specific recombination. *EMBO J.* **18**, 5724–5734 (1999).
52. Galitski, T. & Roth, J. R. Pathways for homologous recombination between chromosomal direct repeats in *Salmonella typhimurium*. *Genetics* **146**, 751–767 (1997).
53. Haber, J. E. DNA recombination: the replication connection. *Trends Biochem. Sci.* **24**, 271–275 (1999).
54. Malkova, A., Ivanov, E. L. & Haber, J. E. Double-strand break repair in the absence of RAD51 in yeast: a possible role for break-induced DNA replication. *Proc. Natl Acad. Sci. USA* **93**, 7131–7136 (1996).
55. Morrow, D. M., Connelly, C. & Hieter, P. “Break copy” duplication: a model for chromosome fragment formation in *Saccharomyces cerevisiae*. *Genetics* **147**, 371–382 (1997).
56. Lundblad, V. & Blackburn, E. H. An alternative pathway for yeast telomere maintenance rescues est1-senescence. *Cell* **73**, 347–360 (1993).
57. Le, S., Moore, J. K., Haber, J. E. & Greider, C. W. RAD50 and RAD51 define two pathways that collaborate to maintain telomeres in the absence of telomerase. *Genetics* **152**, 143–152 (1999).
58. Griffith, J. D. *et al.* Mammalian telomeres end in a large duplex loop. *Cell* **97**, 503–514 (1999).
59. Bosco, G. & Haber, J. E. Chromosome break-induced DNA replication leads to nonreciprocal translocations and telomere capture. *Genetics* **150**, 1037–1047 (1998).
60. Chen, C., Umez, K. & Kolodner, R. D. Chromosomal rearrangements occur in *S. cerevisiae* rfa1 mutator mutants due to mutagenic lesions processed by double-strand-break repair. *Mol. Cell* **2**, 9–22 (1998).
61. Cordeiro-Stone, M., Makhov, A. M., Zaritskaya, L. S. & Griffith, J. D. Analysis of DNA replication forks encountering a pyrimidine dimer in the template to the leading strand. *J. Mol. Biol.* **289**, 1207–1218 (1999).
62. Opperman, T., Murli, S., Smith, B. T. & Walker, G. C. A model for a umuDC-dependent prokaryotic DNA damage checkpoint. *Proc. Natl Acad. Sci. USA* **96**, 9218–9223 (1999).

Correspondence and requests for materials should be addressed to M.E.G.

A ribonucleotide reductase gene involved in a p53-dependent cell-cycle checkpoint for DNA damage

Hiroshi Tanaka, Hirofumi Arakawa, Tatsuya Yamaguchi, Kenji Shiraishi, Seisuke Fukuda, Kuniko Matsui, Yoshiki Takei & Yusuke Nakamura

Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan

The p53 gene is frequently inactivated in human cancers. Here we have isolated a p53-inducible gene, *p53R2*, by using differential display to examine messenger RNAs in a cancer-derived human cell line carrying a highly regulated wild-type p53 expression system. *p53R2* contains a p53-binding sequence in intron 1 and encodes a 351-amino-acid peptide with striking similarity to the ribonucleotide reductase small subunit (R2), which is important in DNA synthesis during cell division. Expression of *p53R2*, but not R2, was induced by ultraviolet and γ -irradiation and adriamycin treatment in a wild-type p53-dependent manner. Induction of *p53R2* in p53-deficient cells caused G2/M arrest and prevented cells from death in response to adriamycin. Inhibition of endogenous *p53R2* expression in cells that have an intact p53-dependent DNA damage checkpoint reduced ribonucleotide reductase activity, DNA repair and cell survival after exposure to various genotoxins. Our results indicate that *p53R2* encodes a ribonucleotide reductase that is directly involved in the p53 checkpoint for repair of damaged DNA. The discovery of *p53R2* clarifies a relationship between a ribonucleotide reductase activity involved in repair of damaged DNA and tumour suppression by p53.

The p53 gene is inactivated more frequently than any other gene in a wide range of human cancers¹. In its normal state, it exerts a tumour-suppressing function by regulating cell-cycle arrest or by inducing apoptosis². Its encoded product, a transcription factor, binds to DNA in a sequence-specific manner to activate transcription of target genes such as those for p21/WAF1, MDM2, BAX and GADD45^{2,3}. Among the known targets, p21/WAF1 is a well-known mediator in the p53 signalling pathway that induces cell-cycle arrest at the G1 phase through inhibition of cyclin-dependent kinase^{4,5}. Without functional p21/WAF1, cells expressing wild-type p53 fail to arrest the cell cycle^{6,7}; therefore, a normal complement of target genes appears to be essential for p53 to function effectively.

In addition to its important functions in cell-cycle arrest and apoptosis⁸⁻¹⁰, accumulating evidence has indicated that p53 may be involved in DNA repair caused by various genotoxic stresses; for example, fibroblasts homozygous for mutant p53 are deficient in global DNA repair¹¹. Wild-type p53 protein is required for global genomic nucleotide excision repair in ultraviolet-induced DNA damage¹² and protects cells from death after irradiation^{13,14}. Inactivation of p53 protein enhances sensitivity to multiple chemotherapeutic drugs^{15,16}.

The identification of genes regulated by p53 is an important step towards understanding these physiological functions. Here we report a new ribonucleotide reductase gene, designated *p53R2*, that is a direct target for p53. We also demonstrate the p53-dependent DNA repair mechanism by which p53 can regulate DNA repair activity after genotoxic stresses through transcriptional induction of *p53R2*.

The ribonucleotide reductase gene, *p53R2*

We previously identified four p53-target genes (*GML*¹⁷, *P2XM*¹⁸, *BAI1*¹⁹ and *CSR*²⁰) through the direct cloning of p53-binding sequences from human genomic DNA²¹. We have also applied differential display²² methods to isolate p53 target genes, using a human cell line in which expression of exogenous wild-type p53 can be regulated under the control of the lactose operon²³. The parental cell line (colon cancer SW480) lacks any wild-type p53, but a cell

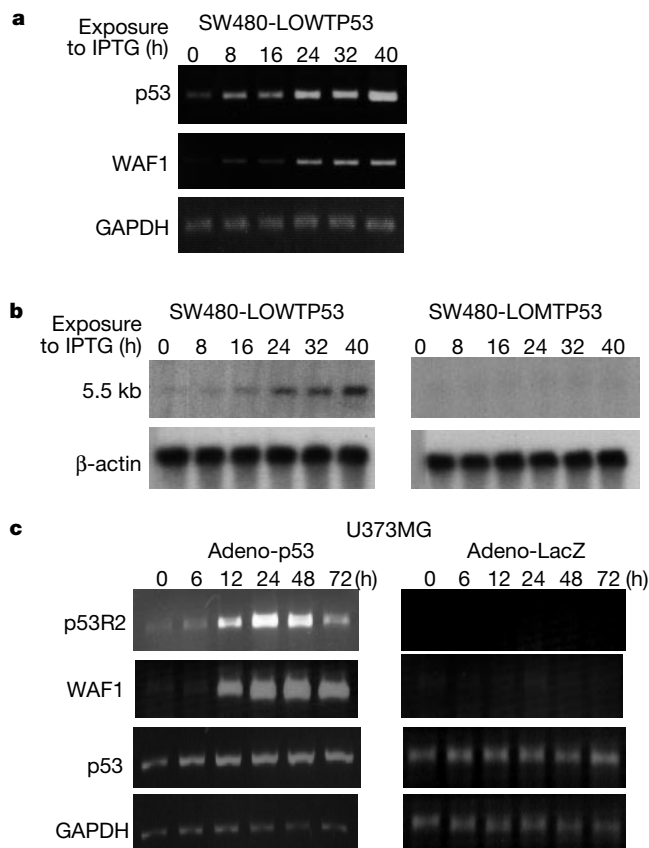


Figure 1 Isolation and characterization of the *p53R2* gene. **a**, Induction of p53 and WAF1 by exposure to IPTG. Expression of the GAPDH gene was examined in the same cells as a quantity control. **b**, Northern blot analysis of *p53R2* induction in SW480-LOWTP53 and SW480-LOMTP53. β -actin expression was used as a quantity control. **c**, Induction of p53, WAF1 and *p53R2* by adeno-p53 or adeno-LacZ. Expression of the GAPDH gene was examined in the same cells as a quantity control.

line established from it (SW480-LOWTP53) can express exogenous (transfected) wild-type *p53* in a time-dependent fashion after exposure to 5 mM isopropylthiogalactoside (IPTG) (Fig. 1a). *p21/WAF1* is also inducible in this derivative cell line following induction of wild-type *p53*. By this approach, we isolated two additional *p53*-inducible genes, *TP53TG1* (ref. 23) and *TP53TG3* (ref. 24), and subsequently detected a 5.5-kilobase (kb) transcript that was also inducible by wild-type *p53* but not mutant *p53* (SW480-LOWTP53) (Fig. 1b). This transcript was also strongly induced by infection of adeno-*p53* but not adeno-LacZ in U373MG cells in a time-dependent manner (Fig. 1c).

Using the 132-base pair (bp) differential display fragment representing this transcript as a probe to screen a skeletal-muscle complementary DNA library, we obtained an almost full-length cDNA sequence consisting of 4,955 nucleotides. This cDNA incorporated an open reading frame encoding a 351-amino-acid peptide with 80% identity to a small subunit (R2) of human ribonucleotide reductase (Fig. 2), which catalyses conversion of ribonucleoside diphosphates to the corresponding deoxyribonucleotides to provide a balanced supply of precursors for DNA synthesis²⁵. Comparison of this sequence with those of yeast *RNR2* (ref. 26) and *RNR4* (refs 27, 28) revealed 60% and 40% identity, respectively (Fig. 2), and we considered this gene to be a human counterpart of yeast *RNR2*. Hence, we called it *p53R2* (for *p53*-inducible ribonucleotide reductase small subunit 2 homologue). The ribonucleotide reductase small subunit signature and two putative nuclear localization signal sequences are shown in Fig. 2.

p53R2 is a direct target for *p53*

Cloning and sequencing of two cosmid clones containing *p53R2* enabled us to determine the genomic structure of the gene and to

find a possible *p53*-binding site of 20 nucleotides in intron 1. *p53R2* consists of 9 exons that span a 30-kb genomic region (Fig. 3a). We mapped *p53R2* to chromosome 8q23.1 by fluorescence *in situ* hybridization (FISH) analysis using the cosmid clones as probes. Using the candidate *p53*-binding DNA sequence, we performed electromobility-shift assays (EMSA) and observed the shift of cells expressing wild-type *p53*. Supershift of this band by addition of anti-*p53* monoclonal antibodies Pab421 and Pab1801 further supported the conclusion that the 20-bp sequence within *p53R2* is a binding site of *p53* (Fig. 3b).

We examined whether this sequence has *p53*-dependent transcriptional activity. The heterogeneous reporter vector (550RE-Luc) was constructed by cloning a 550-bp segment corresponding to intron 1 of *p53R2*, including a potential *p53*-binding sequence upstream of the SV40 minimal promoter. The 550RE-Luc reporter vector was co-transfected into H1299, a *p53*-null lung cancer cell line, with either of the expression plasmids designed to express wild-type *p53* or mutant *p53*. Luciferase activity of 550RE-Luc was increased by co-transfection of the wild-type *p53* expression vector but not that of mutant *p53* (Fig. 3c). Moreover, a point mutation in the potential *p53*-binding sequence in the 550 bp segment of *p53R2* completely impaired this activity (Fig. 3c). Furthermore, we prepared two reporter vectors whose promoter was connected to one (20RE-Luc) or two copies (40RE-Luc) of a synthesized oligonucleotide corresponding to the potential 20-bp *p53*-binding sequence. Wild-type *p53*, but not mutant *p53*, induced more than 20-fold higher luciferase activity in cells co-transfected with 40RE-Luc vector. This result, combined with that of EMSA, clearly indicates that *p53R2* is a direct target gene for *p53*.

We then examined expression of *p53R2* and R2 at each phase

R2	MLSLRVPLAPITDPQQLQSLPLKGLSLVDKENTPPAL	1-37
RNR2	MPKETPSKAAADALSDLEIKDKSKNLNKELETRENNRVKSDM	1-43
<i>p53R2</i>	MGDPERPEAAAGLDQDERSSSDTNESEIKSNEEPLLRKSSRRFVIFPIQY	1-49
R2	SGTRVLASKTARRIFQETPEPKTKAAAPGVEDEPLLENPRRFVIFPIEY	38-87
RNR2	LKEKLSKDAENHKAYLQSHQVHRHKLKEMEKEEPLLNEDKERQVLFPIKY	44-93
RNR4	MEAHNQFLKTFQKERHDMKEAEKDEILLMENSRRFVVFPIKY	1-42
<i>p53R2</i>	PDIWKMYKQAAQASFWTAEVDLSKDLPHW-NKLKADERYFISHILAFFAA	50-98
R2	HDIWQMYKKAQASFWTAEVDLSKDIQHW-ESLKPERRYFISHVLAFFAA	88-136
RNR2	HE-WQAYEASFWTAEAEIDLSKDIHDWNNRMNENERRFFISRVLAFFAA	94-142
RNR4	HE-WAAYRKVEASFWTAEIEIAKDTEDF-QKLTDDQKTYIGNLLALSIS	43-90
1		
<i>p53R2</i>	SDGIVNENLVERFSQEVQVPEARCFYGFQIETENVHSEMYSLIDITYIRD	99-148
R2	SDGIVNENLVERFSQEVQITEARCFYGFQIAMENHSEMYSLIDITYIKD	137-186
RNR2	SDGIVNENLVENFSTEVQIPEAKSFYGFQIMIENTHSETYSLLIDITYIKD	143-192
RNR4	SDNLVKNKYLIENTFSAQLQNPEGKSFYGFQIMMENIYSEVYSMMVDAFFKD	91-140
2		
<i>p53R2</i>	PKNREFLFNAIETMPVKKKADWALRWIADRKSTFGERVVAFAAVEGVFF	149-198
R2	PKREFLFNAIETMPVKKKADWALRWICDKEATYGERVVAFAAVEGVFF	187-236
RNR2	PKSEFLFNIAHTIPEIGEKAEWALRWIQADALFGERLVAFASIEGVFF	193-242
RNR4	PKNIP-LFKEIIANLPEVVKHKAFFIERWISNDDSLYAEERLVAFAAKEGVFF	141-189
3		
<i>p53R2</i>	SGSFAAIFWLKKRGLMPGLTFSNELISRDEGLHCDFACLMFQYLVNKPSE	199-248
R2	SGSFAAIFWLKKRGLMPGLTFSNELISRDEGLHCDFACLMFQYLVNKPSE	237-286
RNR2	SGSFAAIFWLKKRGLMPGLTFSNELISRDEGLHCDFACLMFQYLVNKPSE	243-292
RNR4	AGNYASMFWLTDKKIMPGGLAMANNRNICRDRGAYTDFSCCLFAHLRTKPNP	190-239
<i>p53R2</i>	ERVREIIVDAVKIEQEFLEALPVGLIGMNCILMKQYIEFVADRLLVELG	249-298
R2	ERVREIIVNAVRIEQEFLEALPVGLIGMNCILMKQYIEFVADRLLVELG	287-336
RNR2	AIVEKIIVTAEVEIEQRYFLDALPVALLGMNADLMNQYVEFVADRLLVAFG	293-342
RNR4	KIIEKIIVTAEVEIEQRYFLDALPVALLGMNADLMNQYVEFVADRLLVAFG	240-289
<i>p53R2</i>	FSKVFOAENPFDFMENISLECKTNFFFEKRVSEYQRFVAVM--AETT---DN	299-343
R2	FSKVFRVENPFDFMENISLECKTNFFFEKRVSEYQRFVAVM--SSPT---EN	337-381
RNR2	NKKYKVENPFDFMENISLECKTNFFFEKRVSDYQKAGVM--SKSTKQEG	343-390
RNR4	NEKYNAVNPFFDFMENISLECKTNFFFEKRVSDYQKASDMSKSATP---SK	290-336
<i>p53R2</i>	VFTLDADF	344-351
R2	SFTLDADF	382-389
RNR2	AFTFNEDF	391-398
RNR4	EINFDDDF	337-344

Figure 2 Comparison of the amino-acid sequences of *p53R2*, R2, RNR2 and RNR4. Identities are indicated in black. The ribonucleotide reductase small subunit signature is indicated by region 1, and nuclear localization signals by regions 2 and 3.

a

p53 CBS
RRRRCWWGYYYYRRRCWWGYYYY
tGACATGCCCCAGGCATGTCT
— 1kb

BS

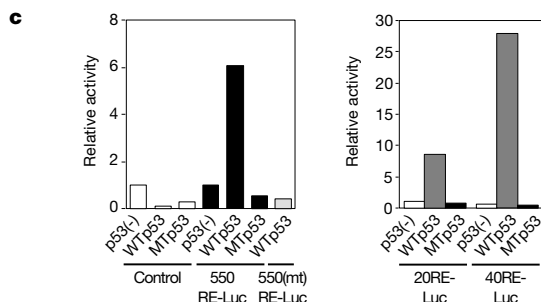
E1 E2 E3 E4 E5 E6 E7 E8 E9

b

p53	+	+	+	+	+	+	+	+
Pab421	-	-	+	+	+	+	-	+
Pab1801	-	-	-	-	-	-	-	+
Competitor DNA (TL)	-	+	-	+	-	-	-	-
Competitor DNA (self)	-	-	-	-	+	-	-	-



p53R2-p53BS PC-p53BS



Ectopic expression of *p53R2* in p53-deficient cells

To elucidate whether the arrest was due to ribonucleotide reductase activity by ectopic overproduction of p53R2, ribonucleotide reductase inhibitor (HU) was added to p53R2-expressing H1299 cells 48 h after they were treated with adriamycin. Addition of HU prevented G2/M arrest and resulted in cell death (Fig. 5a, Myc-p53R2 Ad(+)HU(+) and p53R2 Ad(+)HU(+)), as in the parental cells (Fig. 5a, vector). Moreover, translocation and accumulation of p53R2 protein occurred in the nuclei of cells in which DNA damage had been induced by adriamycin (Fig. 5b, Myc-p53R2 Ad(+)), whereas staining in cytoplasm and nucleus was diffuse and weak in cells without DNA damage (Fig. 5b, Myc-p53R2 Ad(-)). DNA damage in NHDF cells caused G1 and G2 arrest (Fig. 5a,

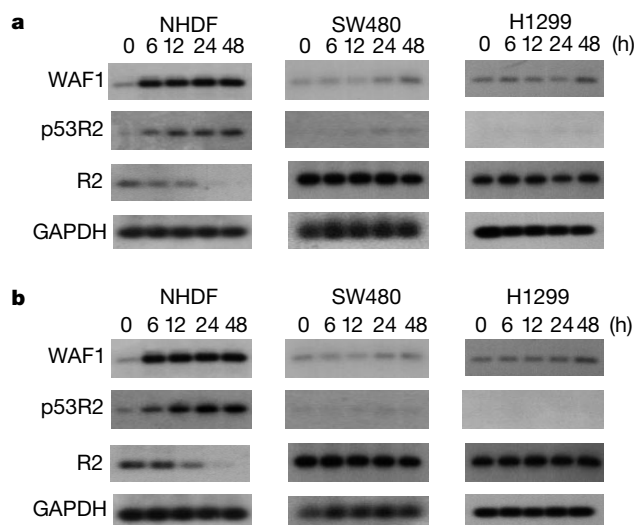


Figure 4 p53-dependent induction of *p53R2* transcription in human cells after DNA damage by 14 Gy γ -radiation (**a**) or 0.2 $\mu\text{g mL}^{-1}$ adriamycin (**b**). Expression of the GAPDH gene was used as a quantity control in RT-PCR experiments to measure induction of p53R2, p21/WAF1 and R2 expression after γ -irradiation. NHDF, a normal human dermal fibroblast cell line, contains wild-type p53 genes. SW480 (a colorectal adenocarcinoma cell line) and H1299 (a lung carcinoma cell line) both lack wild-type p53.

NHDF Ad(+)), and the cells sustained those arrests after HU treatment (Fig. 5a, NHDF Ad(+)/HU(+)). The results suggest that ectopic expression of p53R2 alone in a p53-deficient cell line can restore G2/M arrest and cell survival against DNA damage, through ribonucleotide reductase activity.

Induction of p53R2 and DNA repair activity by genotoxins

To investigate DNA repair through ribonucleotide reductase activity of p53R2, we used a DNA synthesis assay coupled with measurement of ribonucleotide reductase activity by tracing reduction of [³H]rCDP and uptake of its derivative [³H]dCTP into a damaged genome in permeabilized cells after DNA damage using a modified published method³². To elucidate the role of endogenous p53R2 for p53-dependent DNA repair activity, we used MCF7 cells with wild-type p53 protein and p53-dependent DNA damage checkpoint function. As shown in Fig. 6a, after DNA damage by γ -irradiation, the expression of p53R2 mRNA was strongly induced, and that of R2 was repressed in a time-dependent manner as observed in NHDF cells. The cell cycle of MCF7 cells was completely blocked at G1 and G2 phases by 12 h, and these blocks were sustained until at least 48 h after DNA damage (Fig. 6c). Because p53R2 may be important for DNA synthesis in the repair process of damaged DNA, we suspected that DNA repair activity could be upregulated in cells arrested at the G1 and G2 phases, and so assayed for DNA repair after γ -irradiation. In damaged MCF7 cells, DNA repair activity through ribonucleotide reductase was markedly enhanced in a time-dependent manner until 48 h after DNA damage (Fig. 6b). p53R2 protein was also strongly induced and accumulated in the nuclei of damaged cells (Fig. 6d, e). This increased activity was inhibited by

HU but not by the addition of dCTP to the reaction mixture (data not shown). The increased activity was not due to DNA synthesis or DNA replication in cells at S phase because the FACS profile clearly showed that cell cycle was completely blocked at the G1 and G2 phase, and almost no cell fraction at S phase was present (Fig. 6c). Moreover, little induction of this activity occurred in H1299, p53-null cancer cells (Fig. 6b). These results indicate that G1 and G2 cell-cycle arrests, expression and nuclear accumulation of p53R2, and activation of ribonucleotide reductase activity were induced for the repair of DNA damage in a p53-dependent manner.

p53 regulates DNA repair through p53R2

To assess the role of p53R2 in DNA damage repair under physiological conditions, we synthesized antisense oligonucleotides, (AS1, AS2 and AS3) and sense oligonucleotides (SE1, SE2 and SE3), corresponding to the p53R2 coding sequence (see Methods). To examine the effect of presence or absence of p53 on induction of p53R2, DNA repair activity and cell survival against various causes of DNA damage, we also synthesized antisense oligonucleotide p53AS, which suppresses p53 protein¹⁶. Fluorescein isothiocyanate-labelled AS3 transfected by lipofectin was efficiently incorporated into the nuclei of transfected MCF7 cells 4 h after transfection (Fig. 7a), and suppressed induction of endogenous p53R2 protein 48 h and 72 h after DNA damage (Fig. 7b). p53AS also markedly suppressed induction of p53R2, as well as p53 protein, 48 h and 72 h after DNA damage (Fig. 7b), suggesting that induction of p53R2 by DNA damage is clearly dependent on p53 protein. Using AS3, SE3 and p53AS, we examined whether inhibition of p53R2 or p53 could reduce the ribonucleotide reduc-

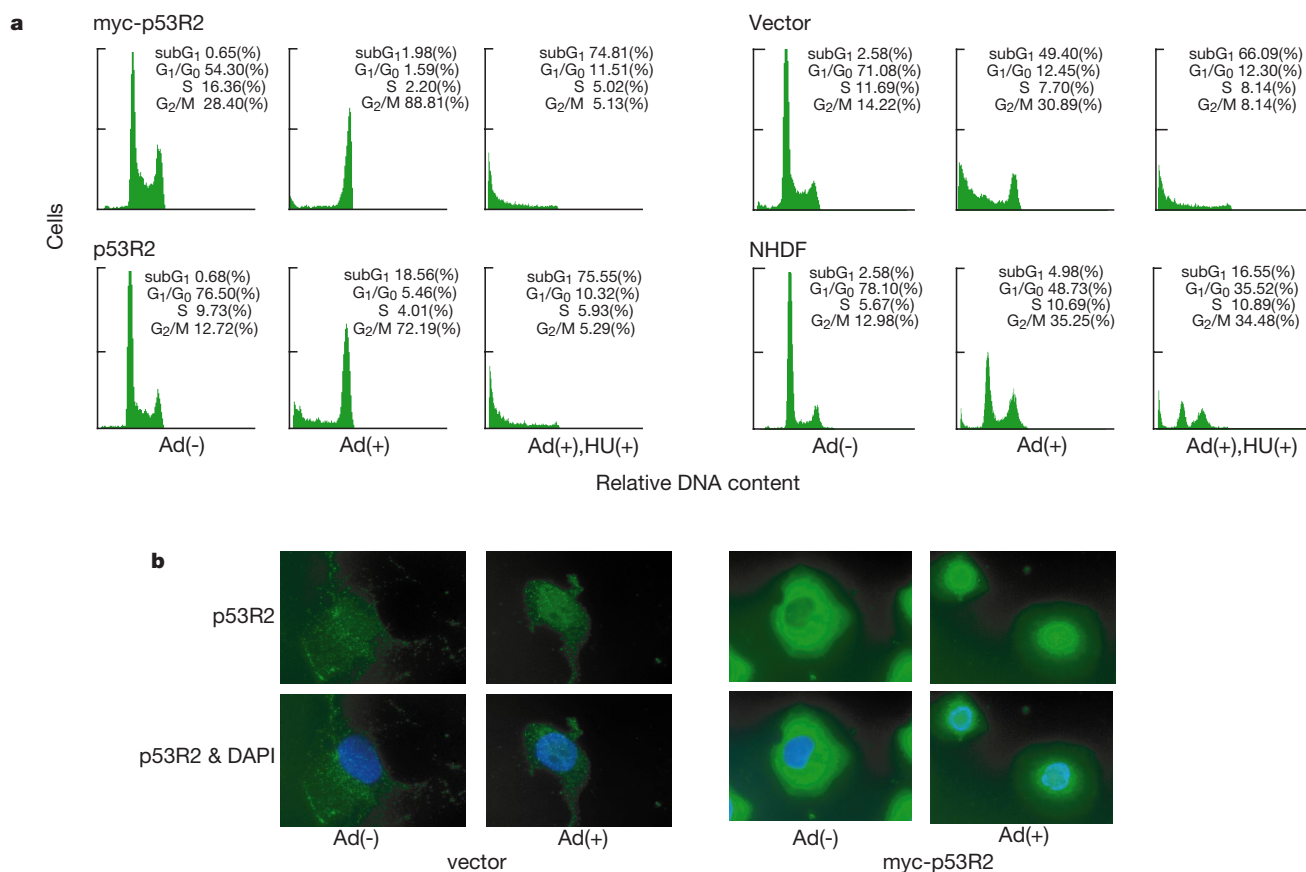


Figure 5 Ectopic expression of p53R2 in p53-deficient cancer cells restores G2/M arrest and cell survival after DNA damage. **a**, Flow-cytometric analysis of p53R2-expressing cells. NHDF and H1299 cells transfected with p53R2, Myc-p53R2 or vector alone were treated with adriamycin (Ad), or with adriamycin and ribonucleotide reductase inhibitor

(HU) added 48 h after the initial addition of adriamycin. Cells treated with adriamycin, or those with adriamycin and HU were then analysed by FACScan 48 h and 72 h after the initial addition of adriamycin, respectively. **b**, Subcellular localization of p53R2 protein in Myc-tagged p53R2-expressing cells 48 h after DNA was damaged with adriamycin.

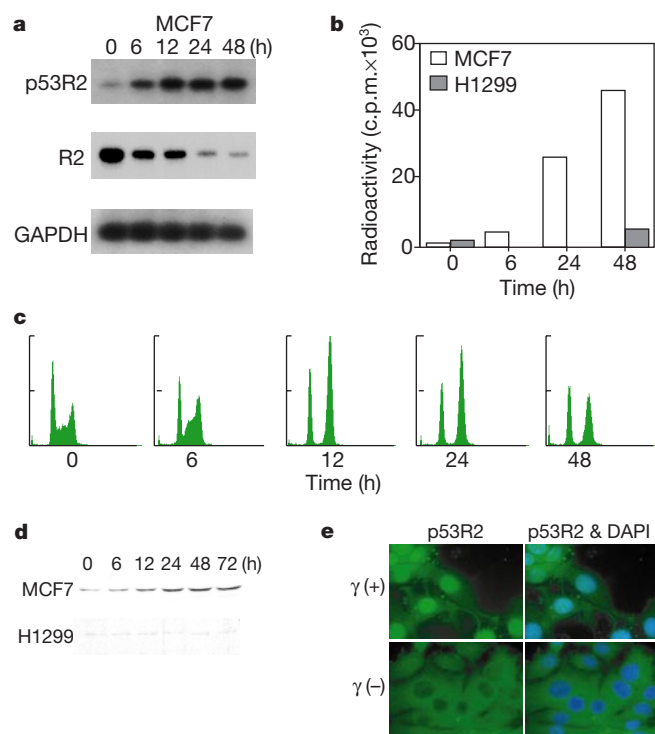


Figure 6 Cell-cycle arrest, p53R2 expression and DNA repair activity through p53R2 ribonucleotide reductase activity after DNA damage. **a**, Induction of *p53R2* mRNA expression and repression of *R2* mRNA expression by DNA damage caused by γ -irradiation in a time-dependent manner. **b**, DNA repair activity of MCF7 and H1299 cells. Uptake of [3 H]dCTP converted from [3 H]rCDP into the damaged genome at various times after γ -irradiation is shown. **c**, FACS analysis of MCF7 cells at various times after γ -irradiation. **d**, Induction of endogenous p53R2 protein by DNA damage caused by γ -irradiation in a time-dependent manner in MCF7 and H1299 cells. **e**, Nuclear accumulation of endogenous p53R2 protein in response to genotoxic stress induced by γ -irradiation. DNA repair assays were repeated at least three times in duplicate samples, and the average score is shown.

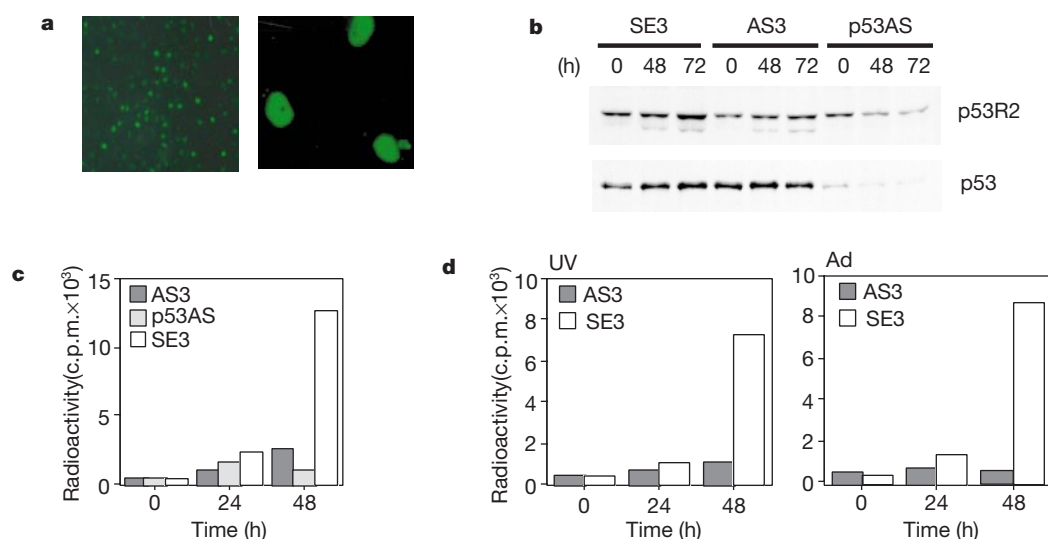


Figure 7 Inhibition of either p53 or p53R2 protein expression reduces DNA repair activity through ribonucleotide reductase after various genotoxic stresses. **a**, Nuclear incorporation of FITC-labelled antisense oligonucleotide AS3 in MCF7 cells at 4 h after transfection. **b**, Inhibition of p53 and p53R2 protein expression by p53AS and AS3 in MCF7 cells at 24 h after DNA damage by γ -irradiation. **c**, Suppression of DNA repair activity by inhibition of either p53 or p53R2 expression in MCF7 cells. Uptake of [3 H]dCTP

converted from [3 H]rCDP into the damaged genome at various times after γ -irradiation is shown. **d**, Suppression of DNA repair activity by inhibition of p53R2 expression in MCF7 cells. Uptake of [3 H]dCTP converted from [3 H]rCDP into the damaged genome at various times after ultraviolet irradiation (UV) or adriamycin (Ad) treatment is shown. DNA repair assays were repeated at least three times in duplicate samples, and the average score is shown.

tase DNA repair activity that occurs after DNA damage. As shown in Fig. 7c, pretreatment of MCF7 cells with either AS3 or p53AS, but not SE3, apparently suppressed incorporation of [3 H]dCTP into the damaged genome, suggesting that the increased DNA repair activity through ribonucleotide reductase was due to p53R2 protein endogenously induced by p53. The similar results were also obtained in MCF7 cells damaged by ultraviolet radiation and adriamycin (Fig. 7d).

We then examined whether dysfunction of DNA repair activity through inhibition of p53R2 expression results in an increase of cell mortality by treatment of various DNA damages, using both the short-term cell survival assay and the long-term colony formation assay. Figure 8 clearly shows that inhibition of p53R2 expression in MCF7 cells significantly reduced survival of cells after exposure to various genotoxins in the short-term cell survival assay and that it also reduced cell survival against genotoxins such as ultraviolet radiation and adriamycin in the colony formation assay (Fig. 8b). Notably, inhibition of p53 expression revealed a similar effect on cell survival in both cell survival assays.

The results presented here clearly indicate that p53R2 has a pivotal role in the repair of DNA damage caused by exposure to various genotoxins. We conclude that p53 is likely to be important in DNA repair through the transcriptional induction of p53R2 in response to DNA damages.

Discussion

Ribonucleotide reductase is a rate-limiting enzyme for DNA synthesis as it catalyses the conversion of ribonucleoside diphosphates to their corresponding deoxyribonucleotides. Evidence from experiments in yeast and mouse has indicated that this enzyme is important in DNA replication and in DNA repair^{23,33}. The significant sequence similarity between p53R2 and the small subunit R2 of ribonucleotide reductase, and the fact that expression of the p53R2 gene is inducible by DNA damage imply that this ribonucleotide reductase has a crucial role in supplying precursors for DNA repair. Indeed, inhibition of p53R2 expression reduced ribonucleotide reductase activity after DNA damage.

Four ribonucleotide reductase subunits (RNR1, RNR2, RNR3

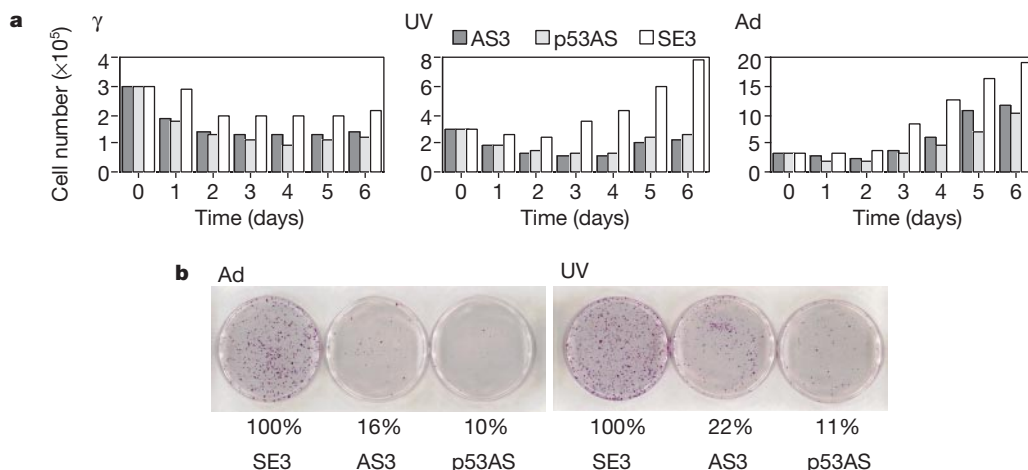


Figure 8 Inhibition of either p53 or p53R2 protein expression reduces cell survival after genotoxic stress. **a**, Increase in cell death in response to ultraviolet (UV) and γ -irradiation and adriamycin (Ad) treatment by inhibition of either p53 or p53R2 expression in MCF7 cells at the indicated times after genotoxic stress in the short-term cell survival assay.

b, Increase in cell death in response to ultraviolet irradiation and adriamycin treatment by inhibition of either p53 or p53R2 expression in MCF7 cells at 3 weeks after genotoxic stress in colony formation assay. Cell survival assays were repeated at least three times in duplicate samples, and the average score is shown.

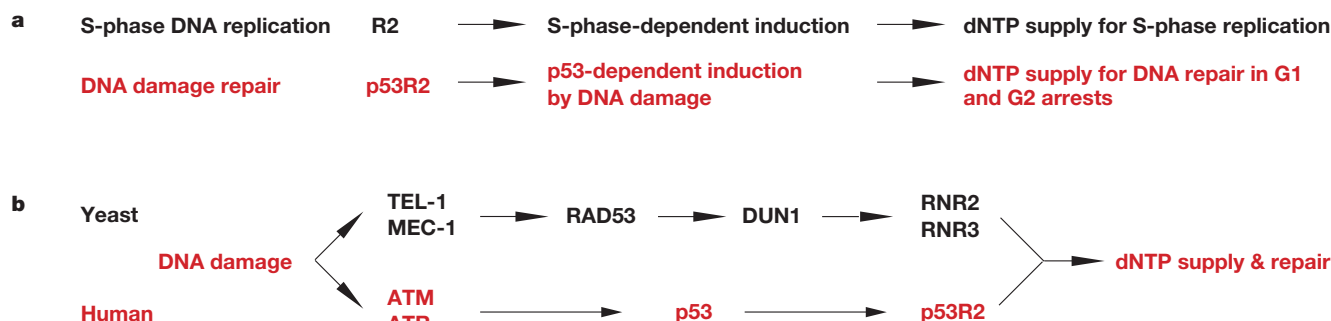


Figure 9 Hypothetical models for DNA damage repair and replication. **a**, The two different roles of ribonucleotide reductases R2 and p53R2. **b**, DNA damage-signalling pathways in

human and yeast.

and RNR4) have been reported in yeast. RNR1³⁴ and RNR3³⁴ correspond to large subunits with 80% identity to each other; RNR2²⁶ and RNR4^{27,28} correspond to a small subunit although RNR4 is lacking dNDP catalysis domain. RNR1 appears throughout the cell cycle and its expression is slightly induced by DNA damage, whereas RNR3 mRNA can be increased more than 100-fold by DNA damage³⁴, indicating that at least two different mechanisms of transcriptional regulation are represented by the large subunits (RNR1 and RNR3). Sequence comparison indicated that p53R2 is a human counterpart of RNR2 as well as of R2 (Fig. 2). However, we have shown that in the human, the transcription-regulating activities of two small units of ribonucleotide reductase, R2 and p53R2, are quite different. One type (yeast RNR1 and human R2) is involved in maintenance dNTPs for DNA replication during S phase, and the other (yeast RNR3 and p53R2) supplies dNTPs for urgent repair of DNA damage (Fig. 9a). p53R2 is likely to supply urgent precursors for DNA synthesis at the actual site of DNA repair—within the damaged nuclei of arrested cells—and those precursors for DNA repair seem to be supplied from ribonucleoside diphosphate reduction by p53R2, and not from the dNTP pool, for the following reasons. First, unlike other known ribonucleotide reductase proteins, p53R2 is translocated into nuclei when cells are arrested by DNA damage. Second, the level of ribonucleotide reductase activity of cells without DNA damage is much less than that of cells with DNA damage. Last, increase of ribonucleotide reductase activity and DNA synthesis after DNA damage is not inhibited by the addition of dCTP to the reaction. We speculate that

the ribonucleotide reductase activity of p53R2 is important not for maintenance of a dNTP pool, but for the urgent supply of precursors from ribonucleoside diphosphates for DNA repair, and that p53R2 may be directly involved in protein complexes of DNA repair machinery. This notion is in contrast to previous reports that R2 is located only in the cytoplasm, and that dNTP molecules are supplied to the nucleus from there^{35,36} to maintain the pool of dNTPs.

In yeast, TEL1 and Mec1, homologues of human ATM and ATR³⁷, activate RAD53 to induce RNR genes³⁸, probably through activation of Dun1³⁹. On the other hand, in the human, ATM activates p53 protein^{40,41}, which then induces p53R2 by directly activating its transcription (Fig. 9b). Our results clearly indicate that p53 could participate in DNA repair activity by direct induction of p53R2 in response to DNA damages. The discovery of p53R2 clarifies that there is a relationship between the ribonucleotide reductase activity involved in repair of DNA damage and the tumour-suppressing function of p53. That is, p53 is directly involved in the ribonucleotide reductase signalling pathway in humans, and ribonucleotide reductase activity may be important for preventing initiation of neoplastic transformation by supplying critical precursors for DNA synthesis from ribonucleoside diphosphates to repair damaged DNA. We propose that inactivation of p53 directly interferes with the transcription of p53R2 in response to DNA damage, with the result that ribonucleotide reductase activity is insufficient for normal DNA repair. Faulty regulation of p53R2 might also enhance mis-incorporation of dNTPs and dysregulation of DNA repair

machinery and thereby increase the frequency of mutations. In other words, the genomic instability one often sees in tumours lacking wild-type p53 may reflect dysfunction of ribonucleotide reductase due to the failure of p53R2 induction. □

Methods

Cell culture

We used a normal human fibroblast cell line derived from neonatal skin (NHDF4042; Clonetic Inc). Human cancer-cell lines SW480 (colorectal adenocarcinoma) and H1299 (lung carcinoma) were from ATCC. All cells were cultured under conditions recommended by their respective depositors.

Semiquantitative RT-PCR and northern blotting

Each RT-PCR experiment was carried out using cDNA generated from 0.2 µg of total RNA. The RT-PCR exponential phase was determined from 15 to 25 cycles to allow semiquantitative comparisons among cDNAs developed from identical reactions. All reactions involved an initial denaturation at 94 °C for 2 min followed by 15–25 cycles at 94 °C for 30 s, 55–59 °C 1 min and 72 °C for 1 min, on a Gene Amp PCR system 9600 (Perkin Elmer). A 2 µg aliquot of each poly(A)⁺ RNA was separated on a 1% agarose gel containing 1 × MOPS buffer and 2% formaldehyde and transferred to nylon membrane. The hybridization with a random-primed ³²P-labelled p53R2 cDNA probe was carried out according to the manufacturer's instructions.

Cell-damage experiment

Cells were subjected to γ-radiation (14 Gy) or ultraviolet (10 J m⁻²) or treated with adriamycin (0.2 µg ml⁻¹). Total RNAs were isolated from the cells 0, 6, 12, 24 and 48 h after treatment. We performed semiquantitative RT-PCR experiments to examine expression of p53R2 and R2 in the damaged cells.

Selection of stable transformants

H1299 cells (1 × 10⁵) were plated in 6-well culture dishes 24 h before transfection of 1 µg of expression vector (pcDNA3.1⁺-Myc-tagged p53R2, p53R2, or vector only) using LipofectAmine Plus reagent (Gibco/BRL) according to the manufacturer's instructions. Stable transformants were selected in the presence of G418 (Gibco/BRL). Resistant clones were expanded and tested for p53R2 by PCR (p53R2 transformants) or western blotting using anti-Myc antibody (for Myc-tagged p53R2 transformants). At least 50 colonies were tested in each case.

Immunocytochemistry

Myc-p53R2 stable transformants were plated on polylysine-coated, multiwell chamber slides and 24 h later the medium was replaced with medium containing 0.2 µg ml⁻¹ of adriamycin. The cells were fixed 48 h later with PBS containing 4% paraformaldehyde and rendered permeable by treatment for 2.5 min at 4 °C with PBS containing 0.1% Triton X-100. The cells were covered with blocking solution (3% BSA and 2.5% goat serum in PBS) for 1 h at room temperature and incubated for 2 h at room temperature with a mouse anti-Myc antibody (PharMingen) diluted 1:1000 in blocking solution. This antibody was stained with a goat anti-mouse secondary antibody conjugated to FITC and viewed with an ECLIPSE E800 microscope (Nikon).

Electromobility-shift assay (EMSA)

Lung-cancer cell line H1299 was infected with an adeno-p53 vector designed to express wild-type p53. Nuclear extracts from these cells were incubated for 30 min at room temperature with 0.5 µg of sonicated salmon sperm DNA, EMSA buffer, the appropriate ³²P-labelled double-stranded (annealed) oligomer and, in some samples, with monoclonal anti-p53 antibodies, Pab421 (Oncogene Science) and/or Pab1801 (Santa Cruz Biotechnology). After incubation, each sample was electrophoresed in a native 4% polyacrylamide gel using 0.5 × TBE buffer. The gels were dried and exposed for autoradiography at -80 °C for 3 h.

Cell-cycle analysis

Cells were trypsinized, washed, collected and fixed with 70% ethanol at the indicated time. Fixed samples were centrifuged treated with RNase (1 mg ml⁻¹), and resuspended in propidium iodide (50 µg ml⁻¹). The stained cells were analysed on a Becton Dickinson FACScan flow cytometer.

Luciferase assay

A 550-bp DNA fragment including a potential p53-binding sequence, TGACATGCCCA-GGCATGTCT, was PCR amplified using primer F (CCGACGCGTCAGGACTCAGTAGA-GGAGC) and primer R (CCGCTCGAGCACAGGGTCAGTGACACTG) and cloned into pGL3-promoter vectors (550RE-Luc; Promega). Oligomer pairs, 1F (CGCGTTGACAT-GCCCGGCATGTCTC) and 1R (TCGAGAGACATGCTGGGCATGTCAA) corresponding to one copy of the potential p53-binding sequence, or 2F (CGCGTTGACATGCCAGGCATGTCTTACATGCCAGGCATGTCTC) and 2R (TCGAGAGACATGCCCTGGGCA TGTCAGACATGCCTGGGCATGTCAA) corresponding to two copies of potential p53-binding sequence, were annealed and cloned into pGL3-promoter vector digested with *Mlu*I and *Xho*I (termed 20RE-Luc and 40RE-Luc, respectively). To make 550(mt)RE-Luc vector, a point mutation 'T' was inserted into the site of the seventh nucleotide 'G' of BS using the QuickChange site-directed mutagenesis

kit (Stratagene). The whole procedure was done according to the manufacturer's instruction (Promega). Quantification of both luciferase activities and calculation of relative ratios were carried out manually with a luminometer.

Antisense oligonucleotides

To inhibit expression of endogenous p53R2, we prepared HPLC-purified antisense oligonucleotides (AS1, TCGCCCATCGCGCAGA; AS2, TTTCCGGTTCGCCCAT; AS3, ACATTTACCTCATCCT) and as controls, sense oligonucleotides (SE1, TCTGCGCATGGGCGA; SE2, ATGGGCGACCCGGAAA; SE3, AGGATGAGGTAA-TGT) according to the sequence of the p53R2 gene. FITC-labelled AS3 was used for evaluation of nuclear incorporation efficiency. Antisense oligonucleotide for p53 (p53AS; CCCTGCTCCCCCTGGCTCC) was synthesized as reported¹⁶.

Cellular viability

Exponentially growing cells were transfected with either sense, SE3, or antisense oligonucleotides, AS3, incubated for 4 h and exposed to DNA damaging agents (10 J m⁻² of ultraviolet radiation, 14 Gy of γ-radiation or 0.2 µg ml⁻¹ of adriamycin for 2 h, and immediately harvested by trypsin. The damaged cells were diluted in 6-well dishes at 3 × 10⁵ cells per well, or in 10-cm dishes at three concentrations, 1 × 10⁵, 1 × 10⁴ and 1 × 10³ cells per dish, for the short-term cell survival assay and colony formation assay, respectively, and maintained with fresh medium. At various time points, cells were harvested and assessed for viability by staining with 0.4% trypan blue for the short-term assay.

DNA repair assay

We evaluated DNA repair by a DNA synthesis assay coupled with measurement of reduction of ribonucleoside diphosphate³² with some modifications. For the DNA-damage experiment with antisense or sense oligonucleotide treatment, 2 × 10⁶ cells were plated on a 10-cm dish, and oligonucleotides (1 µM) were transfected with Lipofectin reagent (Gibco/BRL) the following day. After a 4-h incubation at 37 °C, cells were damaged by 14 Gy of γ-radiation, 10 J m⁻² of ultraviolet 0.2 µg ml⁻¹ of adriamycin for 2 h, and then the medium was changed. The DNA repair assay was done at the indicated times after DNA damage. To standardize the activity, we counted the number of live cells by trypan blue staining immediately after harvesting cells, and used 5 × 10⁵ cells in each experiment to quantify DNA repair activity. For permeabilization, the cells with or without DNA damage were washed twice in solution A (150 mM sucrose, 80 mM KCl, 35 mM HEPES pH 7.4, 5 mM potassium phosphate pH 7.4, 5 mM MgCl₂, 0.5 mM CaCl₂), suspended in cold solution A including 0.25 mg ml⁻¹ lysolecithin (Sigma) and incubated for 1 min at 4 °C. To measure the activity of ribonucleoside diphosphate reduction and DNA synthesis, 5 × 10⁵ permeabilized cells were incubated in 300 µl of 50 mM HEPES pH 7.4, 10 mM MgCl₂, 8 mM dithiothreitol, 0.06 mM FeCl₃, 7.5 mM potassium phosphate pH 7.4, 0.75 mM CaCl₂, 10 mM phosphoenolpyruvate, 0.2 mM [³H]rCDP, 0.2 mM rGDP, 0.2 mM rADP and 0.2 mM dTDP at 37 °C for 10 min. Then 300-µl aliquots of the incubation mixture were added to 60 µl of 60% perchloric acid/0.1% sodium pyrophosphate and kept on ice for 15 min. This mixture was diluted by adding 1 ml of distilled H₂O, and centrifuged to precipitate acid-insoluble material. The pellet was extracted with 0.1 ml of 0.2 N NaOH and incubated at 37 °C for 30 min. This whole process of perchloric acid precipitation and NaOH extraction was repeated, and 75 µl aliquots of the NaOH suspension in 5 ml of AQUASOL-2 (PACKARD) were counted for radioactivity with a Aloka LSC-5100 liquid scintillation counter.

Antibodies

Polyclonal antibodies against p53R2 protein were generated in rabbits using recombinant p53R2 protein synthesized in bacteria. They were affinity purified using the same recombinant protein. Mouse monoclonal anti-p53 antibody (DOO1) was from Oncogene and mouse monoclonal anti-human c-Myc antibody (9E10) was from PharMingen. Each antibody was used for western blot analysis and immunofluorescence staining analysis.

Received 2 November 1999; accepted 6 January 2000.

- Greenblatt, M. S., Bennett, W. P., Hollstein, M. & Harris, C. C. Mutations in the p53 tumor suppressor gene: clues to cancer etiology and molecular pathogenesis. *Cancer Res.* **54**, 4855–4878 (1994).
- el-Deiry, W. S., Kern, S. E., Pietenpol, J. A., Kinzler, K. W. & Vogelstein, B. Definition of a consensus binding site for p53. *Nature Genet.* **1**, 45–49 (1992).
- Levine, A. J. p53, the cellular gatekeeper for growth and division. *Cell* **88**, 323–331 (1997).
- el-Deiry, W. S. et al. WAF1, a potential mediator of p53 tumor suppression. *Cell* **75**, 817–825 (1993).
- Harper, J. W., Adami, G. R., Wei, N., Keyomarsi, K. & Elledge, S. J. The p21 cdk-interacting protein cip1 is a potent inhibitor of G1 cyclin-dependent kinases. *Cell* **75**, 805–816 (1993).
- Deng, C., Zhang, P., Harper, J. W., Elledge, S. J. & Leder, P. Mice lacking p21CIP1/WAF1 undergo normal development, but are defective in G1 checkpoint control. *Cell* **82**, 675–684 (1995).
- Brugarolas, J. et al. Radiation-induced cell cycle arrest compromised by p21 deficiency. *Nature* **377**, 552–557 (1995).
- Kastan, M. B., Onyekwere, O., Sidransky, D., Vogelstein, B. & Craig, R. W. Participation of p53 protein in the cellular response to DNA damage. *Cancer Res.* **51**, 6304–6311 (1991).
- Kuerbitz, S. J., Plunkett, B. S., Walsh, W. V. & Kastan, M. B. Wild-type p53 is a cell cycle checkpoint determinant following irradiation. *Proc. Natl Acad. Sci. USA* **89**, 7491–7495 (1992).
- Lane, D. P. Cancer. p53, guardian of the genome. *Nature* **358**, 15–16 (1992).
- Ford, J. M. & Hanawalt, P. C. Li-Fraumeni syndrome fibroblasts homozygous for p53 mutations are deficient in global DNA repair but exhibit normal transcription-coupled repair and enhanced UV resistance. *Proc. Natl Acad. Sci. USA* **92**, 8876–8880 (1995).
- Ford, J. M. & Hanawalt, P. C. Expression of wild-type p53 is required for efficient global genomic nucleotide excision repair in UV-irradiated human fibroblasts. *J. Biol. Chem.* **272**, 28073–28080 (1997).

13. Smith, M. L. & Fornace, A. J. Jr p53-mediated protective responses to UV irradiation. *Proc. Natl Acad. Sci. USA* **94**, 12255–12257 (1997).
14. Eller, M. S., Maeda, T., Magnoni, C., Atwal, D. & Gilchrist, B. A. Enhancement of DNA repair in human skin cells by thymidine dinucleotides: evidence for a p53-mediated mammalian SOS response. *Proc. Natl Acad. Sci. USA* **94**, 12627–12632 (1997).
15. Hawkins, D. S., Demers, G. W. & Galloway, D. A. Inactivation of p53 enhances sensitivity to multiple chemotherapeutic agents. *Cancer Res.* **56**, 892–898 (1996).
16. Ceraline, J. *et al.* Inactivation of p53 in normal human cells increases G2/M arrest and sensitivity to DNA-damaging agents. *Int. J. Cancer* **75**, 432–438 (1998).
17. Furuhashi, T., Tokino, T., Urano, T. & Nakamura, Y. Isolation of a novel GPI-anchored gene specifically regulated by p53; correlation between its expression and anti-cancer drug sensitivity. *Oncogene* **13**, 1965–1970 (1996).
18. Urano, T. *et al.* Cloning of P2XM, a novel human P2X receptor gene regulated by p53. *Cancer Res.* **57**, 3281–3287 (1997).
19. Nishimori, H. *et al.* A novel brain-specific p53-target gene, BAI1, containing thrombospondin type 1 repeats inhibits experimental angiogenesis. *Oncogene* **15**, 2145–2150 (1997).
20. Han, H. J., Tokino, T. & Nakamura, Y. CSR, a scavenger receptor-like protein with a protective role against cellular damage caused by UV irradiation and oxidative stress. *Hum. Mol. Genet.* **7**, 1039–1046 (1998).
21. Tokino, T. *et al.* p53 tagged sites from human genomic DNA. *Hum. Mol. Genet.* **3**, 1537–1542 (1994).
22. Liang, P. & Pardee, A. B. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* **257**, 967–971 (1992).
23. Takei, Y., Ishikawa, S., Tokino, T., Muto, T. & Nakamura, Y. Isolation of a novel TP53 target gene from a colon cancer cell line carrying a highly regulated wild-type TP 53 expression system. *Genes Chromosomes Cancer* **23**, 1–9 (1998).
24. Ng, C. C. *et al.* Isolation and characterization of a novel TP53-inducible gene, TP53TG3. *Genes Chromosomes Cancer* **26**, 329–335 (1999).
25. Reichard, P. From RNA to DNA, why so many ribonucleotide reductases? *Science* **260**, 1773–1777 (1993).
26. Elledge, S. J. & Davis, R. W. Identification and isolation of the gene encoding the small subunit of ribonucleotide reductase from *Saccharomyces cerevisiae*: a DNA damage-inducible gene required for mitotic viability. *Mol. Cell. Biol.* **7**, 2783–2793 (1987).
27. Huang, M. & Elledge, S. J. Identification of RNR4, encoding a second essential small subunit of ribonucleotide reductase in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **17**, 6105–6113 (1997).
28. Wang, P. J. *et al.* Rnr4p, a novel ribonucleotide reductase small-subunit protein. *Mol. Cell. Biol.* **17**, 6114–6121 (1997).
29. Eriksson, S. & Martin, D. W. Jr Ribonucleotide reductase in cultured mouse lymphoma cells. Cell cycle-dependent variation in the activity of subunit protein M2. *J. Biol. Chem.* **256**, 9436–9440 (1981).
30. Eriksson, S., Graslund, A., Skog, S., Thelander, L. & Tribukait, B. Cell cycle-dependent regulation of mammalian ribonucleotide reductase. The S phase-correlated increase in subunit M2 is regulated by *de novo* protein synthesis. *J. Biol. Chem.* **259**, 11695–11700 (1984).
31. Engstrom, Y. *et al.* Cell cycle-dependent expression of mammalian ribonucleotide reductase. Differential regulation of the two subunits. *J. Biol. Chem.* **260**, 9114–9116 (1985).
32. Reddy, G. P. & Pardee, A. B. Coupled ribonucleoside diphosphate reduction, channeling, and incorporation into DNA of mammalian cells. *J. Biol. Chem.* **257**, 12526–12531 (1982).
33. Wright, J. A. *et al.* Regulation and drug resistance mechanisms of mammalian ribonucleotide reductase, and the significance to DNA synthesis. *Biochem. Cell. Biol.* **68**, 1364–1371 (1990).
34. Elledge, S. J. & Davis, R. W. Two genes differentially regulated in the cell cycle and by DNA-damaging agents encode alternative regulatory subunits of ribonucleotide reductase. *Genes Dev.* **4**, 740–751 (1990).
35. Engstrom, Y., Rozell, B., Hansson, H. A., Stemme, S. & Thelander, L. Localization of ribonucleotide reductase in mammalian cells. *EMBO J.* **3**, 863–867 (1984).
36. Engstrom, Y. & Rozell, B. Immunocytochemical evidence for the cytoplasmic localization and differential expression during the cell cycle of the M1 and M2 subunits of mammalian ribonucleotide reductase. *EMBO J.* **7**, 1615–1620 (1988).
37. Elledge, S. J. Cell cycle checkpoints: preventing an identity crisis. *Science* **274**, 1664–1672 (1996).
38. Weinert, T. DNA damage and checkpoint pathways: molecular anatomy and interactions with repair. *Cell* **94**, 555–558 (1998).
39. Zhou, Z. & Elledge, S. J. DUN1 encodes a protein kinase that controls the DNA damage response in yeast. *Cell* **75**, 1119–1127 (1993).
40. Banin, S. *et al.* Enhanced phosphorylation of p53 by ATM in response to DNA damage. *Science* **281**, 1674–1677 (1998).
41. Canman, C. E. *et al.* Activation of the ATM kinase by ionizing radiation and phosphorylation of p53. *Science* **281**, 1677–1679 (1998).

Acknowledgements

This work was supported in part by the 'Research for the Future' Program of the Japan Society for the Promotion of Science.

Correspondence and requests for materials should be addressed to Y.N. (e-mail: yusuke@ims.u-tokyo.ac.jp). The DDBJ accession numbers for human p53R2 are AB036063, AB036524–AB036532.

Evidence of atmospheric sulphur in the martian regolith from sulphur isotopes in meteorites

James Farquhar, Joel Savarino, Terri L. Jackson & Mark H. Thiemens

Department of Chemistry-0356, University of California, San Diego, La Jolla, California 92093, USA

Sulphur is abundant at the martian surface, yet its origin and evolution over time remain poorly constrained^{1,2}. This sulphur is likely to have originated in atmospheric chemical reactions, and so should provide records of the evolution of the martian atmosphere, the cycling of sulphur between the atmosphere and crust, and the mobility of sulphur in the martian regolith³⁻⁶. Moreover, the atmospheric deposition of oxidized sulphur species could establish chemical potential gradients in the martian near-surface environment, and so provide a potential energy source for chemolithoautotrophic organisms⁷. Here we present measurements of sulphur isotopes in oxidized and reduced phases from the SNC meteorites—the group of related achondrite meteorites believed to have originated on Mars—together with the results of laboratory photolysis studies of two important martian atmospheric sulphur species (SO₂ and H₂S). The photolysis experiments can account for the observed sulphur-isotope compositions in the SNC meteorites, and so identify a mechanism for producing large abiogenic ³⁴S fractionations in the surface sulphur reservoirs. We conclude that the sulphur data from the SNC meteorites reflects deposition of oxidized sulphur species produced by atmospheric chemical reactions, followed by incorporation, reaction and mobilization of the sulphur within the regolith.

The SNC meteorites are mainly igneous rocks (basalts and

peridotites), that also contain secondary alteration minerals and impact glasses that preserve chemical and isotopic signatures which are interpreted to originate in the martian atmosphere⁸. This signature has been used to provide insights into chemistry within the martian atmosphere⁹. It has been suggested that $\delta^{34}\text{S}$ for sulphides in SNC meteorites may reflect interactions between the martian atmosphere and regolith^{10,11}, but not conclusively. Until now, the strongest evidence that the martian atmosphere has a role in the martian sulphur cycle is the presence of sulphur-enriched fines at both Viking and Pathfinder landing sites^{1,2} and spectroscopic data supporting the presence of sulphate at the martian surface¹². These have been interpreted as a consequence of atmospheric sulphur deposition^{5,6}, but others postulate that the sulphur-enriched fines represent aeolian deposits of sulphur that have hydrothermal or evaporative, rather than atmospheric origin^{3,4,13,14}.

Four sulphur-isotope measurements of oxidized and reduced sulphur in SNC meteorites strongly indicate that atmospheric chemistry contributes to the martian sulphur cycle. Table 1 lists the results of sulphur-isotope analyses. We note the variable $\Delta^{33}\text{S}$ (0.071 to -0.301) relative to that observed for previous meteorite measurements (Fig. 1). Thermodynamic, kinetic and biotic fractionation processes derive from the mass differences between the different sulphur isotopes and produce highly correlated arrays of constant $\Delta^{33}\text{S}$ ($\delta^{33}\text{S} - 1000((1 + \delta^{34}\text{S}/1000)^{0.515} - 1)$) and $\Delta^{36}\text{S}$ ($\delta^{36}\text{S} - 1000(1 + \delta^{34}\text{S}/1000)^{1.90} - 1$). Variable $\Delta^{33}\text{S}$ among the SNC meteorites requires fractionation processes other than purely classical thermodynamic, kinetic, or biotic processes. Unfortunately, large uncertainties for $\Delta^{36}\text{S}$ that arise because of low isotopic abundance coupled with small sample size prevent us from making an interpretation on the basis of ³⁶S. Regression of the SNC sulphur-isotope data gives the relationship $(1 + \delta^{33}\text{S}/1000) = (1 + \delta^{34}\text{S}/1000)^{0.488 \pm 0.003}$.

Several mechanisms can produce mass-independent sulphur-isotope compositions; however, not all of these mechanisms can meet the criteria required by the SNC meteorites. Cosmic-ray

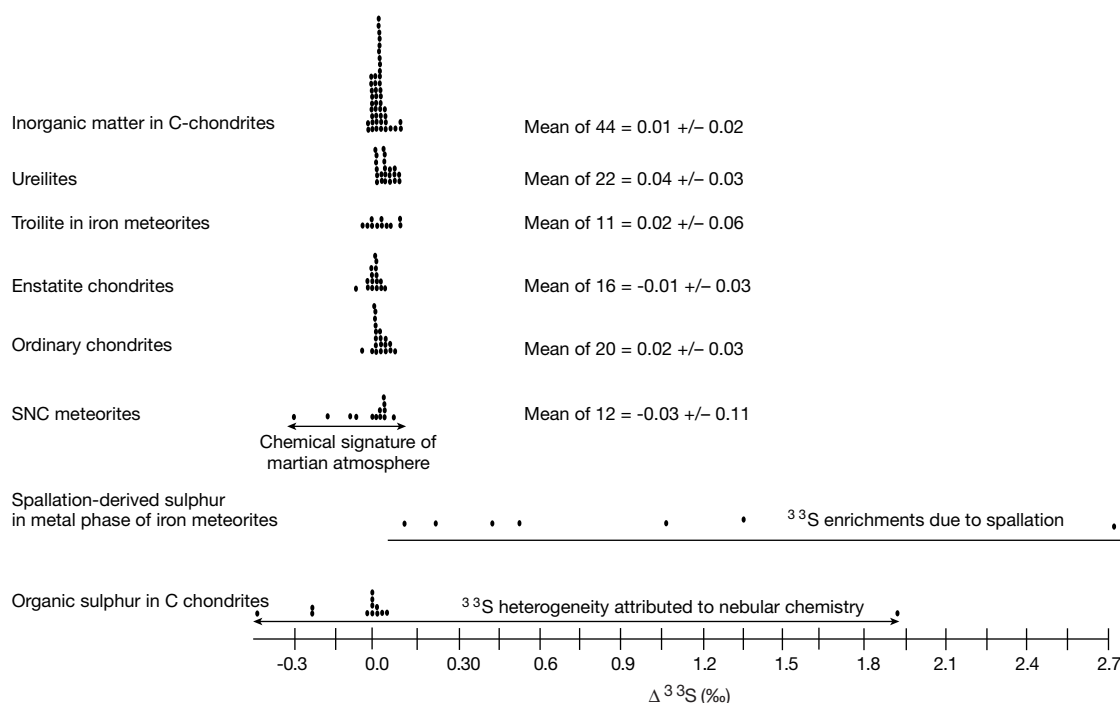


Figure 1 Frequency plot $\Delta^{33}\text{S}$ for SNC meteorites relative to that observed for other meteorites. Data for Ureilites are from UCSD. Meteorite data are from refs 15–18. We note the heterogeneity for $\Delta^{33}\text{S}$ for sulphur extracted from the SNC meteorites. The only other data that exhibit similar heterogeneity are for sulphonic acids extracted from

Murchison¹⁸ and for the metal phase in iron meteorites¹⁵. The former has been interpreted in the context of mass-independent sulphur-isotope chemistry in the presolar nebula¹⁸. The latter has been interpreted in the context of cosmic-ray spallation reactions involving iron¹⁵.

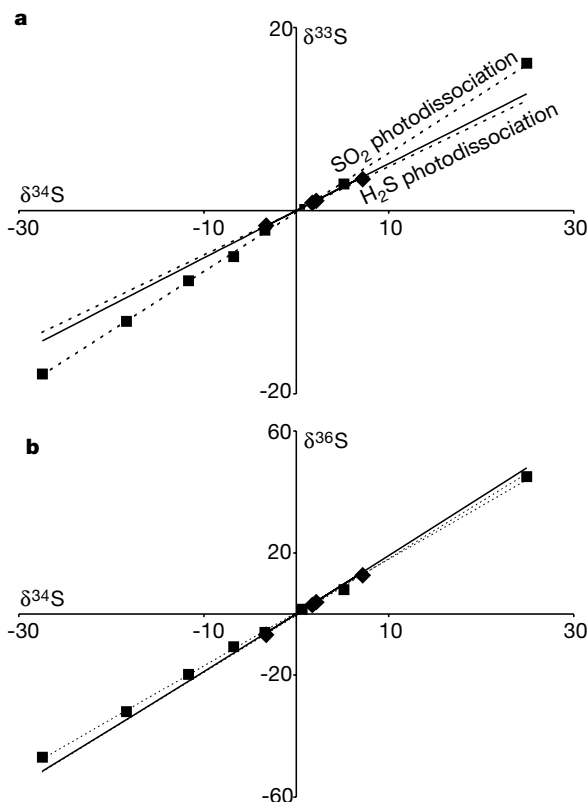


Figure 2 Three isotope plots for photolysis of SO_2 and H_2S . The mass-dependent fractionation lines (solid lines) are included in each plot. Dashed lines are regressed through the data. **a**, $\delta^{33}\text{S}$ versus $\delta^{34}\text{S}$. **b**, $\delta^{36}\text{S}$ versus $\delta^{34}\text{S}$. All experiments were undertaken with a high-pressure Oriel Xe lamp that produces near-continuum deep-ultraviolet radiation starting at ~ 200 nm. Pure H_2S (diamonds) was photolysed at pressures of 10–20 torr. The absorption of deep ultraviolet by H_2S is nearly continuous, starting at about 250 nm with a maximum at 187 nm (ref. 29). Photodissociation of H_2S to HS and H occurs for wavelengths below 317 nm and is followed by production of S due to HS decomposition by several reaction paths²⁹. SO_2 (squares) was photolysed at a variety of conditions including 10–20 torr of pure SO_2 , as a 50/50 mixture of SO_2 and CO_2 with a total pressure of 12 torr, and mixed with 1 atmosphere of nitrogen in the presence of water vapour. The absorption spectrum of SO_2 is complex in the ultraviolet, exhibiting three spectral regions of absorption—extremely weak absorption between 340 and 390 nm, weak absorption between 260 and 340 nm, and strong absorption between 180 and 235 nm (ref. 29). SO_2 exhibits phosphorescent states, several fluorescent states and predissociation states in the spectral region of the near-continuum ultraviolet used in our experiments.

spallation cannot account for the observations because it would produce positive¹⁵, not negative $\Delta^{33}\text{S}$. Likewise, the negative $\Delta^{33}\text{S}$ rules out mass-independent artefacts associated with atmospheric escape. It is also unlikely that heterogeneous accretion or addition of cometary material can account for the sulphur-isotope systematics because highly fractionated reservoirs of primordial sulphur are rare and confined to certain meteoritic organic compounds^{15–18}. Gas-phase chemical reactions generate a number of mass-independent isotopic compositions among species in the terrestrial atmosphere¹⁹ and there is no reason to suppose that this does not occur in the martian atmosphere. In our opinion, the most likely origin of the observed mass-independent sulphur-isotope fractionations is gas phase chemistry. In this context, it is interesting to note that the meteorite that contains phases with the largest ^{17}O excesses (Nakhla)^{20,21} is also the meteorite that possesses sulphur with the largest ^{33}S depletions. These ^{17}O excesses reflect an atmospheric oxygen component in the secondary phases that make up these meteorites.

It is well established that photopolymerization of CS_2 produces large mass-independent fractionations among sulphur isotopes²², but CS_2 is a minor component in the martian atmosphere. The most likely sulphur species in the martian atmosphere (SO_2 and H_2S) derive from volcanic, metamorphic and photochemical sources^{23,24}. These species have short atmospheric lifetimes because they are photochemically active in the optically thin martian atmosphere²⁵. On the basis of atmospheric models for the early Earth²⁶, the martian atmosphere is inferred to have been optically thin even early in martian history. Our photolysis experiments with H_2S and SO_2 produce mass-independent sulphur-isotope compositions (Fig. 2). H_2S photodissociation experiments produced arrays that obey the relationship $(1 + \delta^{33}\text{S}/1000) = (1 + \delta^{34}\text{S}/1000)^{0.485 \pm 0.005}$ and $(1 + \delta^{36}\text{S}/1000) = (1 + \delta^{34}\text{S}/1000)^{1.86 \pm 0.05}$; SO_2 photooxidation experiments produced products that obeyed the relationship $(1 + \delta^{33}\text{S}/1000) = (1 + \delta^{34}\text{S}/1000)^{0.649 \pm 0.006}$ and $(1 + \delta^{36}\text{S}/1000) = (1 + \delta^{34}\text{S}/1000)^{1.75 \pm 0.02}$.

Our SNC sulphur-isotope data provide evidence for a part of the martian sulphur cycle starting with introduction of sulphur-bearing gases to the atmosphere, followed by photochemistry and concluded by incorporation of sulphur into the regolith (Fig. 3). From a planetary perspective, volcanic injections of H_2S and SO_2 into the martian atmosphere are transient point sources²³. This, combined with a short lifetime, will result in heterogeneous distribution of H_2S , SO_2 and their product species in the martian environment, creating a situation conducive for production of isotopically distinct reservoirs of sulphur-bearing species. Cycling of SO_2 to the atmosphere during contact heating of sulphate-rich surface lithologies²³ or by photochemical decomposition of surface

Table 1 SNC sulphur-isotope data

Meteorite	$\delta^{33}\text{S}$	$\delta^{34}\text{S}$	$\delta^{36}\text{S}$	$\Delta^{33}\text{S}$	$\Delta^{36}\text{S}$	s.e. $\Delta^{33}\text{S}$	s.e. $\Delta^{36}\text{S}$
EET 79001.342, 150 °C, H_3PO_4	–0.07	–0.11	0.67	–0.012	0.9	0.004	0.2
EET 79001.518, 25 °C, H_3PO_4	–0.14	–0.33	–0.22	0.026	0.4	0.001	0.4
EET 79001.518, 150 °C, H_3PO_4	–0.43	–0.92	–1.60	0.040	0.2	0.009	0.1
EET 79001.518, 150 °C, H_3PO_4	–0.40	–0.86	–1.52	0.049	0.1	0.032	0.3
Zagami, 25 °C, H_3PO_4	–0.08	–0.16	–0.31	0.001	0.0	0.016	0.2
Shergotty USNM 321*	0.02	–0.01	0.07	0.025	0.1	0.021	0.3
Shergotty USNM 321 sulphate	2.60	5.25	17.95	–0.101	n.d.	0.044	n.d.
Lafayette USNM 1505 sulphate	–2.52	–4.98	–8.86	0.049	0.7	0.009	0.1
Lafayette USNM 1505 pyrite	–1.65	–3.34	–4.3	0.071	2.1	0.051	1.5
Nakhla.23, 150 °C, H_3PO_4	–0.12	0.12	2.47	–0.183	2.2	0.010	0.8
Nakhla USNM 426*	0.41	0.96	4.47	–0.080	2.6	0.012	1.6
Nakhla USNM 426 sulphate	0.74	2.03	6.30	–0.302	2.4	0.026	1.4

* Acid volatile sulphur (monosulphides).

H_3PO_4 acidifications are probably monosulphide. All values are expressed in units of per mil (‰). $\delta^{36}\text{S} = 1000[(^{36}\text{S}/^{32}\text{S})_{\text{sample}}/(^{36}\text{S}/^{32}\text{S})_{\text{reference}} - 1]$ where the reference is Canyon Diablo Troilite for the meteorite sulphur values and the starting composition for the H_2S or SO_2 used in the experiments for the experimental values. Standard errors are given for $\Delta^{33}\text{S}$ and $\Delta^{36}\text{S}$ and are determined from repeat mass spectrometry. Sulphur was extracted from SNC meteorites using two approaches. H_3PO_4 acidifications were undertaken for bulk-rock oxygen-isotope analysis of carbonate at 25 °C and 150 °C (ref. 9). H_2S produced during acidification of these samples was separated by gas chromatography and converted to Ag_2S by reaction with cadmium acetate followed by reaction with silver nitrate solution. Sulphur-isotope ratios of Ag_2S were determined using procedures described in refs 16 and 17. Sulphur-isotope determinations of acid volatile sulphides (monosulphides), sulphate and pyrite were undertaken using sequential extraction procedures^{16,17}. Measurements of $\delta^{34}\text{S}$ are similar to previously measured values for $\delta^{34}\text{S}$ for these meteorites^{10,20}. The sulphur-isotope systematics of monosulphide (for example, FeS) from shergottites provide a baseline for comparison of sulphur-isotope systematics for sulphur in these and other SNC meteorites. The shergottites were EET 79001, Shergotty and Zagami: $\delta^{33}\text{S} = -0.18 \pm 0.19\text{‰}$, $\delta^{34}\text{S} = -0.40 \pm 0.40\text{‰}$ and $\delta^{36}\text{S} = -0.5 \pm 0.9\text{‰}$, and $\Delta^{33}\text{S} = 0.021 \pm 0.019$. ($\Delta^{33}\text{S} = \delta^{33}\text{S} - 1000[(1 + \delta^{34}\text{S}/1000)^{0.515} - 1]$.) s.e., standard error; n.d., not determined.

sulphate²⁴ may also generate SO₂ that could react to form mass-independently fractionated sulphur isotopes in the atmosphere. Constant cycling of sulphur-rich surface dust to the atmosphere by dust devils may play a role in homogenization of surface sulphur. The observed mass-independent compositional variations, however, reflect incomplete homogenization by dust devils and the photochemical decomposition of sulphur-rich dust that is lofted by dust devils may also constitute an important link in cycling between solid-phase and gas-phase sulphur species. Photochemical reaction of gas-phase sulphur-bearing species generated by these processes produces mass-independently fractionated sulphur that can react with surface minerals, or be deposited as product aerosols. These cycles provide a means of generating and transferring a mass-independent isotopic character from the atmosphere to the regolith.

It has been suggested that Nakhla and Lafayette are part of the same stratigraphy²⁷ and that their sulphur-isotope systematics reflect greater interactions with a surface reservoir by fluids affecting Nakhla and lesser, but longer-lived interactions with a surface reservoir by fluids that affected Lafayette¹⁰. Our $\Delta^{33}\text{S}$ measurements of Nakhla and Lafayette are consistent with this interpretation. Nakhla has heterogeneous, but negative $\Delta^{33}\text{S}$, and Lafayette has homogeneous and slightly positive $\Delta^{33}\text{S}$. This observation is also echoed by the higher $\Delta^{17}\text{O}$ of water, carbonate and sulphate in Nakhla relative to Lafayette^{20,21}, which has been interpreted to indicate relative inputs of martian atmospheric oxygen to secondary phases in these samples. These results reflect an active geochemical system where exchange and chemical reactions transport material of atmospheric origin into and within the regolith.

A number of studies have suggested that ^{34}S fractionations can be used to prospect for biological activity on Mars. Large ^{34}S fractionations should not be considered sufficient to indicate biological activity because large ^{34}S fractionations can also be produced by martian atmospheric reactions such as those examined in this study. The variability of $\Delta^{33}\text{S}$ among phases in the same samples (Nakhla and Shergotty) as well as variability of $\Delta^{33}\text{S}$ among related samples (nakhrites and shergottites) is consistent with sulphur-isotope fractionation processes originating in the martian atmosphere. These processes also fractionate ^{34}S , making $^{34}\text{S}/^{32}\text{S}$ an ambiguous marker for biological activity. It is essential, therefore, that future studies aiming to use sulphur-isotope fractionations to search for possible biological activity also consider $^{33}\text{S}/^{32}\text{S}$ and $^{36}\text{S}/^{32}\text{S}$ fractionations. Sulphur species processed by biota should form coherent, mass-dependent fractionation arrays.

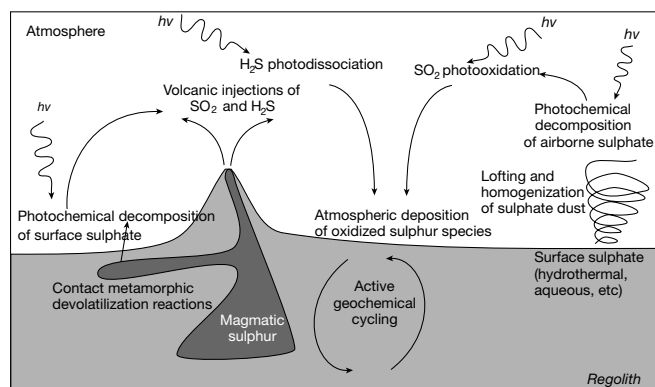


Figure 3 Links in the martian sulphur cycle consistent with the SNC sulphur-isotope data. The short lifetime of H₂S and SO₂ introduced into the martian atmosphere by volcanic injections and escape of gases produced by contact metamorphism²³, as well as SO₂ produced by photochemical decomposition of surface or airborne sulphate²⁴ create a situation where sulphur is cycled through the atmosphere and sulphur isotopes are fractionated by photochemical oxidation processes. Atmospheric deposition of sulphur at the surface, followed by mobilization of this sulphur within the regolith transferred the isotopic signature of this atmospheric chemistry to phases in the SNC meteorites.

Our experiments extend the observation of mass-independent gas-phase chemistry to reactions involving two important sulphur-bearing compounds (H₂S and SO₂) in planetary atmospheres. The signature of these reactions can be used to place additional constraints on the relative inputs of atmospheric sulphur to surface sulphur reservoirs, and subsequent mobility of this sulphur within the regolith with samples of martian regolith, ice and atmosphere. An understanding of this chemistry also bears on our understanding of other planetary atmospheres and its isotopic signature should be sought where sulphur-bearing molecules exist and interact with surface reservoirs such as on Earth, Venus and Io. □

Received 17 September; accepted 9 December 1999.

- Rieder, R. *et al.* The chemical composition of Martian soil and rocks returned by the mobile alpha proton x-ray spectrometer: Preliminary results from the x-ray mode. *Science* **278**, 1771–1774 (1997).
- Clark, B. C. *et al.* Chemical composition of Martian fines. *J. Geophys. Res.* **87**, 10059–10067 (1982).
- Newsom, H. E., Hagerty, J. J. & Goff, F. Mixed hydrothermal fluids and the origin of the Martian soil. *J. Geophys. Res.* **104**, 8717–8728 (1999).
- Toulmin, P. III. *et al.* Geochemical and mineralogical interpretation of the Viking inorganic chemical results. *J. Geophys. Res.* **82**, 4625–4634 (1977).
- Settle, M. & Greeley, R. Formation and deposition of volcanic sulfate aerosols on Mars. *J. Geophys. Res.* **84**, 8343–8354 (1979).
- Banin, A., Han, F. X., Kan, I. & Cicelsky, A. Acidic volatiles and the Mars soil. *J. Geophys. Res.* **102**, 13341–13356 (1997).
- Jakosky, B. M. & Shock, E. L. The biological potential of Mars, the early Earth, and Europa. *J. Geophys. Res.* **103**, 19359–19364 (1998).
- Bogard, D. D. & Johnson, P. Martian gases in an Antarctic meteorite? *Science* **221**, 651–654 (1983).
- Farquhar, J., Thiemens, M. H. & Jackson, T. Atmosphere-surface interactions on Mars: $\Delta^{17}\text{O}$ measurements of carbonate from ALH 84001. *Science* **280**, 1580–1582 (1998).
- Greenwood, J. P., Riciputi, L. R., McSweeney, H. Y., Jr. & Taylor, L. A. Modified sulphur isotopic compositions of sulfides in the nakhrites and Chassigny. *Geochim. Cosmochim. Acta* (in the press).
- Shearer, C. K., Layne, G. D., Papike, J. J. & Spilde, M. N. Sulphur isotopic systematics in alteration assemblages in martian meteorite Allan Hills 84001. *Geochim. Cosmochim. Acta* **60**, 2921–2926 (1996).
- Pollack, J. B. *et al.* Thermal emission spectra of Mars (5.4–10.5 μm) - Evidence for sulfates, carbonates, and hydrates. *J. Geophys. Res.* **95**, 14595–14627 (1990).
- McSweeney, H. Y. & Harvey, R. P. An evaporation model for formation of carbonates in the ALH84001 Martian meteorite. *Int. Geol. Rev.* **40**, 774–783 (1998).
- Warren, P. H. Petrological evidence for low-temperature, possibly flood evaporitic origin of carbonates in the ALH84001 meteorite. *J. Geophys. Res.* **103**, 16759–16773 (1998).
- Gao, X. & Thiemens, M. H. Systematic study of sulphur isotopic composition in iron meteorites and the occurrence of excess ^{33}S and ^{36}S . *Geochim. Cosmochim. Acta* **55**, 2671–2679 (1991).
- Gao, X. & Thiemens, M. H. Isotopic composition and concentration of sulphur in carbonaceous chondrites. *Geochim. Cosmochim. Acta* **57**, 3159–3169 (1993).
- Gao, X. & Thiemens, M. H. Variations of the isotopic composition of sulphur in enstatite and ordinary chondrites. *Geochim. Cosmochim. Acta* **57**, 3171–3176 (1993).
- Cooper, G. W., Thiemens, M. H., Jackson, T. L. & Chang, S. Sulphur and hydrogen isotope anomalies in meteorite sulfonic acids. *Science* **277**, 1072–1074 (1997).
- Thiemens, M. H. Atmosphere science - Mass-independent isotope effects in planetary atmospheres and the early solar system. *Science* **283**, 341–345 (1999).
- Farquhar, J., Thiemens, M. H. & Jackson, T. L. $\Delta^{17}\text{O}$ Anomalies in carbonate from Nakhla and Lafayette and $\Delta^{33}\text{S}$ anomalies in sulphur from Nakhla: Implications for atmospheric chemical interactions with the Martian regolith. *30th Lunar Planet. Sci. Conf.* **30**, 1675 (1999).
- Karlsson, H. R., Clayton, R. N., Gibson, E. K. & Mayeda, T. K. Water in SNC meteorites - evidence for a martian hydrosphere. *Science* **255**, 1409–1411 (1992).
- Colman, J. J., Xu, X. P., Thiemens, M. H. & Troglor, W. C. Photopolymerization and mass-independent sulphur isotope fractionations in carbon disulfide. *Science* **273**, 774–776 (1996).
- Wanke, H. & Dreibus, G. Chemistry and accretion history of Mars. *Phil. Trans R. Soc. Lond. A* **349**, 285–293 (1994).
- Mukhin, L. M., Koscheev, A. P., Dikov, Y. P., Huth, J. & Wanke, H. Experimental simulations of the photodecomposition of carbonates and sulphates on Mars. *Nature* **379**, 141–143 (1996).
- Yung, Y. L., Nair, H. & Gerstell, M. F. CO₂ greenhouse in the early Martian atmosphere: SO₂ inhibits condensation. *Icarus* **130**, 222–224 (1997).
- Kasting, J. F., Pollack, J. B. & Crisp, D. Effects of high CO₂ levels on surface temperature and atmospheric oxidation state of the early Earth. *J. Atmos. Chem.* **1**, 403–428 (1984).
- Harvey, R. P. & McSweeney, H. Y. Petrogenesis of the Nakhla meteorites - evidence from cumulate mineral zoning. *Geochim. Cosmochim. Acta* **56**, 1655–1663 (1992).
- Greenwood, J. P., Riciputi, L. R. & McSweeney, H. Y., Jr. Sulfide isotopic compositions in shergottites and ALH84001, and possible implications for life on Mars. *Geochim. Cosmochim. Acta* **61**, 4449–4453 (1997).
- Okabe, H. *Photochemistry of Small Molecules* (Wiley, New York, 1978).

Acknowledgements

We thank M. Grady, G. McPherson, and M. Lindstrom for access to samples used in this study. We also thank H. McSweeney and H. Newsom for comments and criticisms. This research was supported by the NSF, Calspace and NASA.

Correspondence and requests for materials should be addressed to J. F. (e-mail: jfarquha@ucsd.edu).

Fabrication of photonic crystals for the visible spectrum by holographic lithography

M. Campbell*, D. N. Sharp*, M. T. Harrison*†, R. G. Denning‡ & A. J. Turberfield*

* University of Oxford, Department of Physics, Clarendon Laboratory, Parks Road, Oxford OX1 3PU, UK

‡ University of Oxford, Department of Chemistry, Inorganic Chemistry Laboratory, South Parks Road, Oxford OX1 3QR, UK

The term 'photonics' describes a technology whereby data transmission and processing occurs largely or entirely by means of photons. Photonic crystals are microstructured materials in which the dielectric constant is periodically modulated on a length scale comparable to the desired wavelength of operation. Multiple interference between waves scattered from each unit cell of the structure may open a 'photonic bandgap'—a range of frequencies, analogous to the electronic bandgap of a semiconductor, within which no propagating electromagnetic modes exist^{1–3}. Numerous device principles that exploit this property have been identified^{4–8}. Considerable progress has now been made in constructing two-dimensional structures using conventional lithography³, but the fabrication of three-dimensional photonic crystal structures for the visible spectrum remains a considerable challenge. Here we describe a technique—three-dimensional holographic lithography—that is well suited to the production of three-dimensional structures with sub-micrometre periodicity. With this technique we have made micropatterned polymeric structures, and we have used these as templates to create complementary structures with higher refractive-index contrast.

We generate periodic microstructure by interference of four non-coplanar laser beams in a film of photoresist typically 30 μm thick. The intensity distribution in the interference pattern has three-dimensional translational symmetry; its primitive reciprocal lattice vectors are equal to the differences between the wavevectors of the beams. (Three-dimensional interference patterns have been used to create periodic optical traps for laser-cooled atoms⁹.) Highly exposed photoresist is rendered insoluble; unexposed areas are dissolved away to reveal a three dimensionally periodic structure formed of crosslinked polymer with air-filled voids. The four laser beam wavevectors determine the translational symmetry and lattice constant of the interference pattern; there remain eight parameters, describing the intensities and polarization vectors of the four beams, that are required to define the intensity distribution within a unit cell (that is, the basis of the interference pattern). These parameters allow considerable freedom in determining the distribution of dielectric material within the unit cell, which in turn determines the photonic band structure^{10,11}. (We note that Berger and co-workers have made two-dimensional holographic gratings, as well as proposing, independently of ourselves, a possible extension to three-dimensional lithography¹².)

Figure 1A shows a calculated surface of constant intensity corresponding to a laser interference pattern with face-centred cubic (f.c.c.) symmetry. The wavevectors of the four beams that interfere to produce this pattern are: $\mathbf{k}_0 = 2\pi/d[\frac{3}{2}\frac{3}{2}\frac{3}{2}]$, $\mathbf{k}_1 = 2\pi/d[\frac{5}{2}\frac{1}{2}\frac{1}{2}]$, $\mathbf{k}_2 = 2\pi/d[\frac{1}{2}\frac{5}{2}\frac{1}{2}]$, $\mathbf{k}_3 = 2\pi/d[\frac{1}{2}\frac{1}{2}\frac{5}{2}]$. The difference wavevectors, $\mathbf{k}_1 - \mathbf{k}_0 = 2\pi/d[\frac{1}{2}\frac{1}{2}\frac{1}{2}]$ and so on, generate a body-centred cubic reciprocal lattice corresponding to a real-space f.c.c. interference pattern with a lattice constant $d = 3\sqrt{3}\lambda/2 = 922\text{ nm}$ for a laser wavelength $\lambda = 355\text{ nm}$. This

constant-intensity surface resembles a ball-and-stick model of diamond. The repeated 'abc' pattern of close-packed (111) planes that makes up an f.c.c. lattice is indicated on the side of the cube in Fig. 1A. The primitive basis, repeated at each lattice point, is shown as an inset. It consists of two linked sets of tetrahedral 'bonds' which join it to six nearest neighbours, three each in the (111) planes above and below. The extra-thick central 'bond' along the [111] direction, that links the tetrahedra, corresponds to that joining the two carbon atoms in the diamond basis at (000) and $(\frac{1}{4}\frac{1}{4}\frac{1}{4})$. The connectivity of the structure is identical to that of diamond, and its approximate symmetry is $R\bar{3}m$. We calculate that, by varying the laser intensity or the sensitivity of the photoresist, this interference pattern could be used to create structures with fully connected polymer and air void lattices having filling fractions in the range 34–79%.

To illustrate the flexibility of the holographic process we show in Fig. 1B an equivalent surface of constant intensity, generated by one of many alternative four-beam configurations that can use the same laser wavelength to create distinct f.c.c. lattices. The corresponding laser wavevectors are: $\pi/d[201]$, $\pi/d[\bar{2}01]$, $\pi/d[02\bar{1}]$, $\pi/d[0\bar{2}1]$. In this example the lattice constant is smaller ($d = 397\text{ nm}$) and the basis consists of cuboid 'atoms' connected by 'bonds' to eight of their twelve nearest neighbours.

We use single pulses from an injection-seeded, frequency-tripled, Q-switched Nd:YAG laser (wavelength $\lambda = 355\text{ nm}$) to expose an

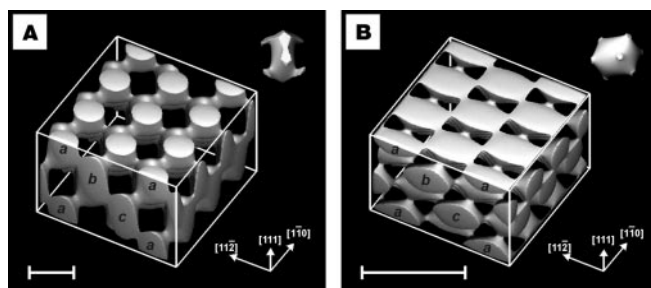


Figure 1 Calculated constant-intensity surfaces in four-beam laser interference patterns designed to produce photonic crystals for the visible spectrum from photoresist. The primitive basis (contents of a Wigner-Seitz unit cell) is shown inset in each case. **A**, f.c.c. pattern with lattice constant 922 nm, used to produce the structures shown in Fig. 3a–d and 4. The close-packed layers of the f.c.c. lattice are indicated on one side of the cube. **B**, f.c.c. pattern with lattice constant 397 nm. Scale bars, 500 nm.

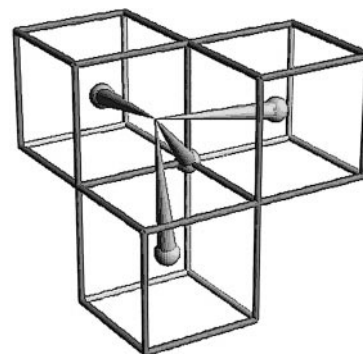


Figure 2 Beam geometry for an f.c.c. interference pattern. The wavevectors of the four laser beams are drawn as cones originating from lattice points in a b.c.c. reciprocal lattice. The differences between the central beam wavevector $\mathbf{k}_0$, which originates from the common point of the three cubes shown, and the three wavevectors $\mathbf{k}_{1-3}$ originating from body-centre lattice points, are the primitive set of reciprocal lattice vectors $2\pi/d(\bar{1}11)$.

† Present address: British Telecom Laboratories, Martlesham Heath, Ipswich IP5 3RE, UK.

epoxy-type photoresist. The four laser beams are generated by splitting the ultraviolet laser output twice with dielectric beam-splitters; their paths to the target are of equal length and contain no lenses or other focusing optics. The beam geometry corresponding to the f.c.c. interference pattern of Fig. 1A is shown schematically in Fig. 2; the three outer beams are symmetrically placed around the central beam, and make an angle of 38.9° with it. The intensity and linear polarization state of each beam are controlled by means of a half-wave plate and a dielectric polarizing beamsplitter; the polarizer is the last optical element before the target.

The photoresist contains the resin Epon-SU8¹³ dissolved in γ -butyrolactone (50–60 wt%) with a triaryl sulphonium salt (0.5–3.0 wt%) acting as a photoacid generator. SU8 has low intrinsic absorption at the laser wavelength, and is capable of sub-0.1- μm resolution¹³. It is also suitable for the production of high-aspect-ratio micromechanical structures formed from ultrathick films¹³. Films of photoresist are spun onto fused silica disks, and the solvent evaporated by heating. The fused silica substrate is index-matched to a thick glass block with hydrocarbon oil to reduce the effects of back-reflected laser light. Absorption of an ultraviolet photon by a molecule of photoacid generator liberates a hydrogen ion; acid-catalysed polymerization does not occur until the film is heated in a post-exposure bake¹³. SU8 has very high functionality with eight epoxy groups per monomer; the network of crosslinks formed is therefore potentially very dense, giving high solubility contrast between strongly and weakly exposed material. The duration of the laser pulse (6 ns) is short compared to the timescales of physical and chemical processes induced by the exposure, so the interference pattern is unperturbed by photoinduced changes in the refractive index of the precursor. The short exposure also eases constraints on the mechanical stability of the optical components. The total dose is adjusted in the range 80–200 mJ cm⁻² to achieve the required

polymer/air ratio in the microperiodic structure. The refractive index of the crosslinked polymer is $n \approx 1.6$. The final step is dissolution of weakly exposed material in propylene glycol methyl ether acetate in an ultrasonic bath. Using this method we have prepared microperiodic films with thicknesses in the range 10–60 μm , corresponding to 14–80 close-packed layers.

Figure 3a shows a scanning electron microscopy (SEM) image of a polymeric microstructure produced by exposure to the interference pattern of Fig. 1A. Refraction at the film surface changes the incident wavevectors, stretching the interference pattern in the [111] direction. During processing, film shrinkage of 10–20% occurs. The developed film is hard and brittle; its top surface is a (111) plane and it has been fractured to reveal (11 $\bar{1}$) cleavage planes. Figure 3b is a magnified view of the (111) surface. The cleavage planes, enlarged in Fig. 3c, reveal the elongated shape of the primitive basis (matching the inset in Fig. 1A).

The surface of a polymeric film that was originally in contact with the substrate is shown as a composite of three overlapping SEM images in Fig. 4A. The optically flat substrate was misaligned by 5° from the (111) plane; the broad bands marked 'abca' that run across the figure correspond to regions where the substrate interface cuts through each close-packed layer. The SEM image therefore contains a large number of parallel cross-sections through the unit cell at slightly different heights that can be stacked to reconstruct the three-dimensional structure shown in Fig. 4B. The structure is very close to the calculated optical intensity contour shown in Fig. 1A. The resultant microstructure, with its non-spherical basis, resembles an optical-frequency version of the microwave photonic crystal of Yablonovitch and co-workers¹⁰. The filling fraction, estimated from the reconstruction in Fig. 4B, is 54%.

To demonstrate that our microstructure has adequate connectivity to act as a template for a crystal with higher refractive-index contrast^{14–18}, we have produced an inverse replica in titania (Fig. 3d) by filling the pores with titanium(IV) ethoxide, and then heating to 575 $^\circ\text{C}$ in moist air to burn off the template and sinter the ceramic. The large ($\sim 30\%$) shrinkage on calcination can be attributed to the major change in titanium density between the tetra-ethoxide and the anatase phase of TiO_2 ($\sim \times 10$); it causes extensive fracturing into small randomly oriented fragments. As a result the material is white, not opalescent; strong scattering across the visible spectrum results from distortions in the templated microstructure, from the random particle orientation, and from inhomogeneities due to the microcrystallinity of the titania. It is known, however, that such shrinkage is lessened if titanium alkoxide is repeatedly (for example, up to

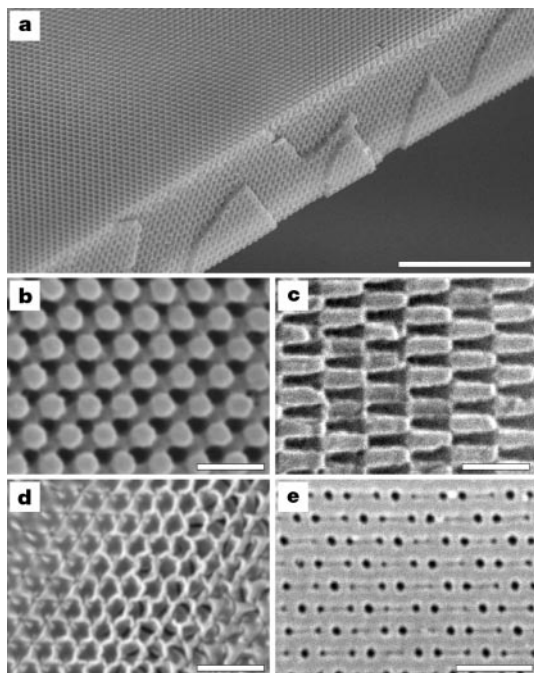


Figure 3 SEM images of photonic crystals generated by holographic lithography. **a**, Polymeric photonic crystal generated by exposure of a 10- μm film of photoresist to the interference pattern shown in Fig. 1A. The top surface is a (111) plane; the film has been fractured along the (11 $\bar{1}$) cleavage planes. Scale bar, 10 μm . **b**, Close-up of a (111) surface. Scale bar, 1 μm . **c**, Close-up of a (11 $\bar{1}$) surface. Scale bar, 1 μm . **d**, Inverse replica in titania made by using the polymeric structure as a template. The surface is slightly tilted from the (111) plane. Scale bar, 1 μm . **e**, (102) surface of a b.c.c. polymeric photonic crystal. Scale bar, 1 μm .

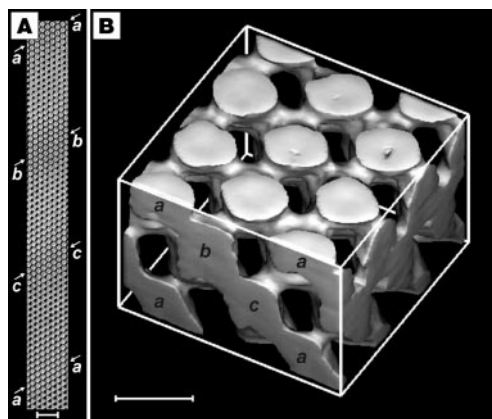


Figure 4 Three-dimensional reconstruction of the surface of an f.c.c. photonic crystal. **A**, SEM image of the bottom surface of a polymeric photonic crystal film that was initially in contact with the fused silica substrate. The surface is at an angle of $\sim 5^\circ$ to the (111) plane. Arrows mark regions where the surface cuts through the close-packed planes of the f.c.c. structure. Scale bar, 2 μm . **B**, The reconstructed surface of the photonic crystal. Scale bar, 500 nm.

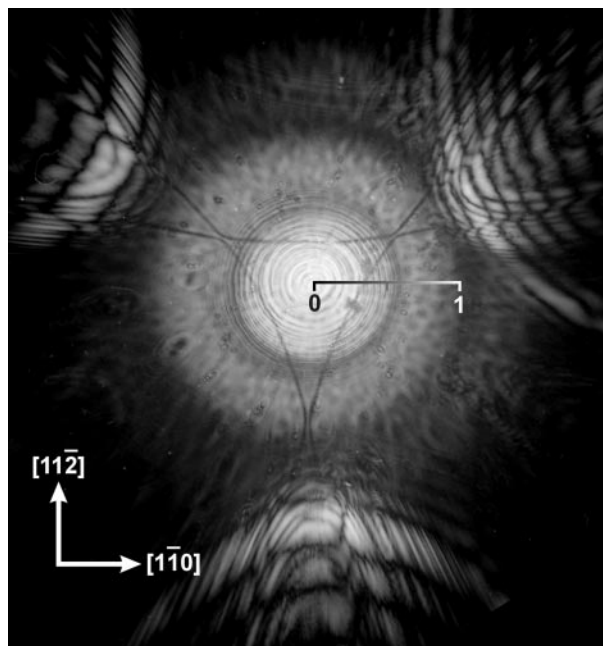


Figure 5 Angular dependence of the intensity of monochromatic light (wavelength, 633 nm) transmitted by the polymeric photonic crystal shown in Fig. 4. Dark and bright Kossel lines identify directions satisfying the Bragg scattering condition in which the transmitted intensity is depleted or enhanced by diffraction by the three-dimensional

photonic crystal. The axis labels correspond to $\tan \alpha$, where α is the angle between the propagation direction and the $[111]$ normal to the film. Axis indicators correspond to the in-plane component of the photon wavevector.

eight times) infiltrated into the interstices of polystyrene colloidal crystals¹⁶, and can be largely eliminated by incorporating nanoparticulate titania directly from a suspension, rather than relying on sol-gel hydrolysis¹⁹. As compacted nanocrystalline ceramic particles can produce macroscopically transparent materials when sintered²⁰, structures with negligible Rayleigh scattering appear feasible. Other successful templating methods, showing less susceptibility to shrinkage distortion, include the electrochemical deposition of semiconductors²¹, the sedimentation of nanoparticulate CdSe (ref. 22), and the inclusion of colloidal gold²³.

To show that microstructures with different translational symmetry can be produced by this technique, Fig. 3e shows an SEM image of the $(\bar{1}02)$ surface of a polymeric film exposed to a body-centred cubic (b.c.c.) interference pattern generated using laser wavevectors $2\pi/d[102]$, $2\pi/d[20\bar{1}]$, $2\pi/d[012]$, $2\pi/d[0\bar{1}2]$. The minimum feature size is 50 nm, demonstrating that—as a result of the strongly nonlinear threshold relationship between exposure dose and solubility—structural elements much smaller than the holographic wavelength are resolvable.

The threshold exposure for the production of insoluble polymer is determined by the optical fluence, by the number density, absorption cross-section and quantum efficiency of initiator (photoacid generator) molecules, and by the chain-length and branching ratio in the polymerization. Attenuation of the laser fluence by absorption in the photoresist therefore limits the thickness of structures that can be made dimensionally homogeneous by this technique. If required, the attenuation can be reduced by offsetting a smaller initiator concentration with a higher exposure dose. A film containing 1% of photoacid generator exhibits good solubility contrast at a dose of $\sim 100 \text{ mJ cm}^{-2}$, and has a photo-initiator transmission coefficient of 99.8% per micrometre at 355 nm. However, commercial SU8 resin contains an impurity that absorbs at 355 nm, causing a further 10% attenuation in a 30- μm film; this impurity is not intrinsic to the resin, and work to remove it is in progress. Using purified resin and our current photoinitiator concentration and exposure dose, we expect to be able to produce photonic crystal templates up to ~ 50 lattice periods

thick with less than 5% attenuation through the thickness of the film.

SEM examination of our polymeric films shows defect-free crystalline order on length scales of hundreds of lattice constants. The injection-seeded Nd:YAG laser operates in single longitudinal and transverse modes, and its temporal (longitudinal) coherence length ($\sim 2 \text{ m}$)—determined by the pulse width—is effectively infinite. However, the transverse mode quality is limited by thermal lensing in the YAG rods, and by diffraction by intra-cavity apertures. Substantial large-scale inhomogeneities, due to the non-uniform intensity distribution across the 0.5-cm^2 beam, cause differential shrinkage and curvature in the films following development. These inhomogeneities occur on a millimetre scale, and are apparent as spatial variations in the iridescent colours of the films in both transmission and reflection.

The films diffract light strongly. The transmission diffraction pattern of a normally incident laser beam has threefold symmetry: it contains six first-order spots, of which three are in general more intense than the others. This pattern is consistent with diffraction from a two-dimensional grating with the hexagonal symmetry of the surface: in other words, it appears to be dominated by coupling between the incident beam and the surface modes. To investigate the bulk optical properties of the structures, we use an optical fibre butted against the surface of a solvent-filled 'f.c.c.' sample to illuminate a small area with monochromatic light from a HeNe laser ($\lambda = 633 \text{ nm}$). The fibre provides a point source of local illumination over all angles within a cone of half-angle $\sim 45^\circ$. Figure 5 shows the angular distribution of the transmitted light (centred circular fringes are due to diffraction by the fibre aperture). The three dark lines that cross the central diffraction maximum are Kossel lines²⁴, a signature of three-dimensional diffraction. The intensity of light travelling in these specific directions is strongly attenuated by Bragg diffraction from the $\{111\}$ lattice planes, and the corresponding diffracted light appears as bright arcs that lie beyond the central diffraction maximum and are the $\{\bar{1}\bar{1}1\}$ Kossel lines. Patterns recorded with air-filled films also include $\{200\}$ Kossel lines.

Other optical techniques for the production of photonic crystals use a tightly focused scanning laser beam to initiate chemical vapour deposition²⁵ or two-photon photopolymerization²⁶. Structures are produced layer-by-layer, so these techniques are relatively slow; they are currently limited to micrometre length scales. Semiconductor microfabrication^{27–29}, which has the required resolution, is an expensive process at the limit of current technology and has not produced structures more than a few unit cells deep. Colloidal crystals may be used as templates to make submicrometre structures^{14–18}, but the use of close-packed spheres severely restricts the range of lattices that may be produced³⁰ and allows almost no freedom to alter the structure of a unit cell. By contrast, the optical properties of microstructures made by holographic lithography may be optimized by controlling the form of the interference pattern. Our fabrication method has the high spatial resolution required to produce photonic crystals for the visible spectrum, as well as creating the connected air and dielectric networks that are important for the opening of a full bandgap. The process is cheap, rapid and potentially scalable for large-scale production. We have also shown that these polymeric structures may be used as templates for the construction of photonic crystals with higher refractive-index contrast. We believe that this is an enabling technology that should allow the theoretical potential of visible optical photonic crystals to be realized. □

Received 13 September; accepted 20 December 1999.

- Yablonovitch, E. Inhibited spontaneous emission in solid-state physics and electronics. *Phys. Rev. Lett.* **58**, 2059–2062 (1987).
- Joannopoulos, J. D., Villeneuve, P. R. & Fan, S. Photonic crystals: putting a new twist on light. *Nature* **386**, 143–149 (1997).
- Krauss, T. F. & De La Rue, R. M. Photonic crystals in the optical regime—past, present and future. *Prog. Quantum Electron.* **23**, 51–96 (1999).
- Yablonovitch, E. et al. Donor and acceptor modes in photonic band structure. *Phys. Rev. Lett.* **67**, 3380–3383 (1991).
- Lin, S.-Y., Chow, E., Hietala, V., Villeneuve, P. R. & Joannopoulos, J. D. Experimental demonstration of guiding and bending of electromagnetic waves in a photonic crystal. *Science* **282**, 274–276 (1998).
- Foresi, J. S. et al. Photonic-bandgap microcavities in optical waveguides. *Nature* **390**, 143–145 (1997).
- Fan, S., Villeneuve, P. R. & Joannopoulos, J. D. High extraction efficiency of spontaneous emission from slabs of photonic crystals. *Phys. Rev. Lett.* **78**, 3294–3297 (1997).
- John, S. & Wang, J. Quantum optics of localized light in a photonic band gap. *Phys. Rev. B* **43**, 12772–12789 (1991).
- Grynberg, G., Lounis, B., Verkerk, P., Courtois, J.-Y. & Salomon, C. Quantized motion of cold cesium atoms in two- and three-dimensional optical potentials. *Phys. Rev. Lett.* **70**, 2249–2252 (1993).
- Yablonovitch, E., Gmitter, T. J. & Leung, K. M. Photonic band structure: the face-centred-cubic case employing nonspherical atoms. *Phys. Rev. Lett.* **67**, 2295–2298 (1991).
- Ho, K. M., Chan, C. T. & Soukoulis, C. M. Existence of a photonic gap in periodic dielectric structures. *Phys. Rev. Lett.* **65**, 3152–3155 (1990).
- Berger, V., Gauthier-Lafaye, O. & Costard, E. Photonic band gaps and holography. *J. Appl. Phys.* **82**, 60–64 (1997).
- Lee, K. Y. et al. Micromachining applications of a high resolution ultrathick photoresist. *J. Vac. Sci. Technol. B* **13**, 3012–3016 (1995).
- Kapitonov, A. M. et al. Photonic stop band in a three-dimensional $\text{SiO}_2/\text{TiO}_2$ lattice. *Phys. Status Solidi A* **165**, 119–123 (1998).
- Holland, B. T., Blanford, C. F. & Stein, A. Synthesis of macroporous minerals with highly ordered three-dimensional arrays of spheroidal voids. *Science* **281**, 538–540 (1998).
- Wijnhoven, J. E. G. J. & Vos, W. L. Preparation of photonic crystals made of air spheres in titania. *Science* **281**, 802–804 (1998).
- Imhof, A. & Pine, D. J. Ordered macroporous materials by emulsion templating. *Nature* **389**, 948–951 (1997).
- Zakhidov, A. A. et al. Carbon structures with three-dimensional periodicity at optical wavelengths. *Science* **282**, 897–901 (1998).
- Subramanian, G., Manoharan, V. N., Thorne, J. D. & Pine, D. J. Ordered macroporous materials by colloidal assembly: a possible route to photonic bandgap materials. *Adv. Mater.* **11**, 1261–1265 (1999).
- Mayo, M. J. Processing of nanocrystalline ceramics from ultrafine particles. *Int. Mater. Rev.* **41**, 85–115 (1996).
- Klein, J. D. et al. Electrochemical fabrication of cadmium chalcogenide microdiode arrays. *Chem. Mater.* **5**, 902–904 (1993).
- Vlasov, Y. A., Yao, N. & Norris, D. J. Synthesis of photonic crystals for optical wavelengths from semiconductor quantum dots. *Adv. Mater.* **11**, 165–169 (1999).
- Velev, O. D., Tessier, P. M., Lenhoff, A. M. & Kaler, E. W. A class of porous metallic nanostructures. *Nature* **401**, 548 (1999).
- Tarhan, İ. İ. & Watson, G. H. Photonic band structure of fcc colloidal crystals. *Phys. Rev. Lett.* **76**, 315–318 (1996).
- Wanke, M. C., Lehmann, O., Müller, K., Wen, Q. Z. & Stuke, M. Laser rapid prototyping of photonic band-gap microstructures. *Science* **275**, 1284–1286 (1997).
- Cumpston, B. H. et al. Two-photon polymerization initiators for three-dimensional optical data storage and microfabrication. *Nature* **398**, 51–54 (1999).

- Cheng, C. C. et al. New fabrication techniques for high quality photonic crystals. *J. Vac. Sci. Technol. B* **15**, 2764–2767 (1997).
- Noda, S., Yamamoto, N. & Sasaki, A. New realization method for three-dimensional photonic crystal in optical wavelength region. *Jpn J. Appl. Phys.* **35**, L909–L912 (1996).
- Fleming, J. G. & Lin, S.-Y. Three-dimensional photonic crystal with a stop band from 1.35 to 1.95 μm . *Opt. Lett.* **24**, 49–51 (1999).
- van Blaaderen, A., Ruel, R. & Wiltzius, P. Template-directed colloidal crystallization. *Nature* **385**, 321–324 (1997).

Acknowledgements

We thank A. J. Wilkinson and the Department of Materials, University of Oxford, for use of scanning electron microscope facilities. This work was supported by the EPSRC/MOD/DERA Joint Grant Scheme, under the Microstructured Photonic Materials Initiative.

Correspondence should be addressed to A.J.T. (a.turberfield@physics.oxford.ac.uk).

Electrophoretic assembly of colloidal crystals with optically tunable micropatterns

R. C. Hayward, D. A. Saville & I. A. Aksay

Department of Chemical Engineering and Princeton Materials Institute, Princeton University, Princeton, New Jersey 08544-5263, USA

The production of materials with micrometre- and submicrometre-scale patterns is of importance in a range of applications, such as photonic materials^{1,2}, high-density magnetic data storage devices³, microchip reactors⁴ and biosensors⁵. One method of preparing such structures is through the assembly of colloidal particles^{5–10}. Micropatterned colloidal assemblies have been produced with lithographically patterned electrodes^{5,11} or micro-moulds¹². Here we describe a different method that combines the well-known photochemical sensitivity of semiconductors^{13,14} with electric-field-induced assembly^{15,16} to create ordered arrays of micrometre-sized colloidal particles with tunable patterns. We show that light affects the assembly processes, and demonstrate

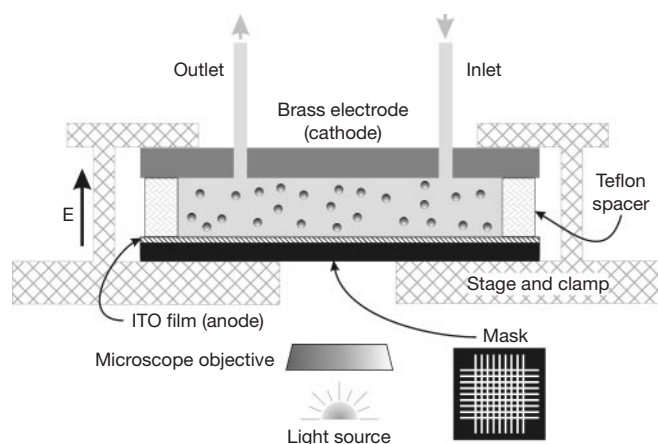


Figure 1 Schematic of the apparatus for particle assembly and pattern formation. The ITO (indium tin oxide) and brass electrodes are separated with a 500- μm Teflon spacer. When desired, the mask is positioned outside the cell between the light source and the glass electrode. Patterns are imaged using a CCD camera mounted on a Leitz Metallovert optical microscope. In most cases, 50 μl of a dilute suspension of polystyrene particles (diameter 2 μm , Duke Scientific) is used to provide a 1- cm^2 lens sandwiched between the electrodes. A positive potential is applied to the ITO electrode while the cathode is grounded; current through the cell is monitored with a multimeter.

how to produce patterns using electrophoretic deposition in the presence of an ultraviolet (UV) illumination motif. The distribution of current across an indium tin oxide (ITO) electrode can be altered by varying the illumination intensity: during the deposition process, this causes colloidal particles to be swept from darkened areas into lighted regions. Illumination also assists in immobilizing the particles on the electrode surface. Although the details of these processes are not well understood, the patterning effects of the UV light are discussed in terms of alterations in the current density^{15,17} that affects particle assembly on an ITO electrode.

We demonstrated the effect of current-density modulation (produced by UV light) on the electrophoretic deposition and assembly of colloid crystals using the apparatus shown in Fig. 1. In most cases, 2- μm polystyrene particles were used; similar observations were made with submicrometre-sized silica. A d.c. electric field strength of $\sim 6,500 \text{ V m}^{-1}$ was used in conjunction with a broad-spectrum light source and a filter that absorbed 100% of the incident radiation at wavelengths below 495 nm. Briefly, our results were as follows. When a potential of 1.3 V was applied with the filter in place, no colloidal crystals formed. On removal of the filter, crystalline aggregates formed rapidly in the illuminated region, and the cell current increased in proportion to the lighted area. When selected regions of the ITO electrode were illuminated in this manner, particles migrated slowly into the illuminated areas. None of the effects of current-density modulation were observed when the ITO

electrode was replaced by a transparent 25-Å layer of sputter-coated iridium, illustrating that a conducting electrode alone is not sufficient for the modulation of colloidal assembly by UV light.

We consider first the process of particle assembly. Assembling colloidal particles by the application of an electric field, a phenomenon first reported by Richetti *et al.*^{18,19}, is a technique for producing “crystalline” particle arrays^{20,21}. Although the details of the assembly mechanism are not fully understood, the process involves simple coulombic interactions which bring particles close to an electrode surface, together with lateral motion stemming from electrohydrodynamic^{15,17} or electroosmotic effects²². Once the particles are close to the surface, where they remain mobile, electrohydrodynamic or electrokinetic processes assemble them into arrays. Particles can be permanently attached to the surface by increasing the attractive force between the particles and the electrode²³. When the attractive forces exceed that due to steric repulsion, entry into the “primary minimum” creates a permanent bond.

In our earlier work^{15,17}, we showed that altering the current density affected the motion of particles and fluid near an electrode. Moreover, it is well-known that photochemical processes alter the current at the interface between a semiconductor and an aqueous solution^{13,14}. Therefore, it follows that illumination patterns could be used to modulate the assembly processes. The experiments described here demonstrate that UV light can be so used. An unusual feature of the technique is that illumination can be externally controlled. Accordingly, the assembly processes can be manipulated or “tuned” by adjusting illumination patterns (and intensity) and applied voltage level (and waveform).

We now consider how the UV light could affect the current density at the ITO electrode. The specificity of this light-effect to the ITO electrode system, and the disappearance of the effect with the introduction of a filter that absorbs light below 495 nm, suggest an interaction between UV light and the electronic band structure of ITO. ITO is a heavily doped n-type semiconductor with a bandgap between 3.5 and 4.3 eV, depending on its composition and method of preparation²⁴. A UV-visible spectrum of the coated electrode used here indicates that it is opaque to light at wavelengths below 310 nm, and transmits between 70 and 90% of radiation above 400 nm. The spectrum agrees well with those reported in the literature^{25,26}.

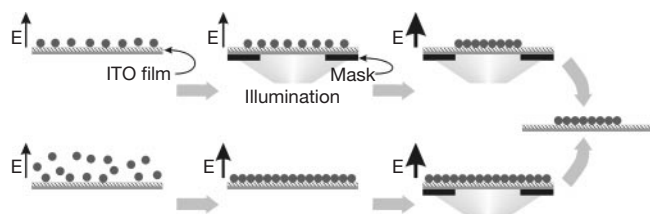


Figure 2 Process schemes for pattern formation. UV light is used to selectively assemble and affix (method A; above) or affix (method B; below) polystyrene particles on the ITO electrode. The thickness of the vertical arrows indicates the (relative) strengths of applied potentials.

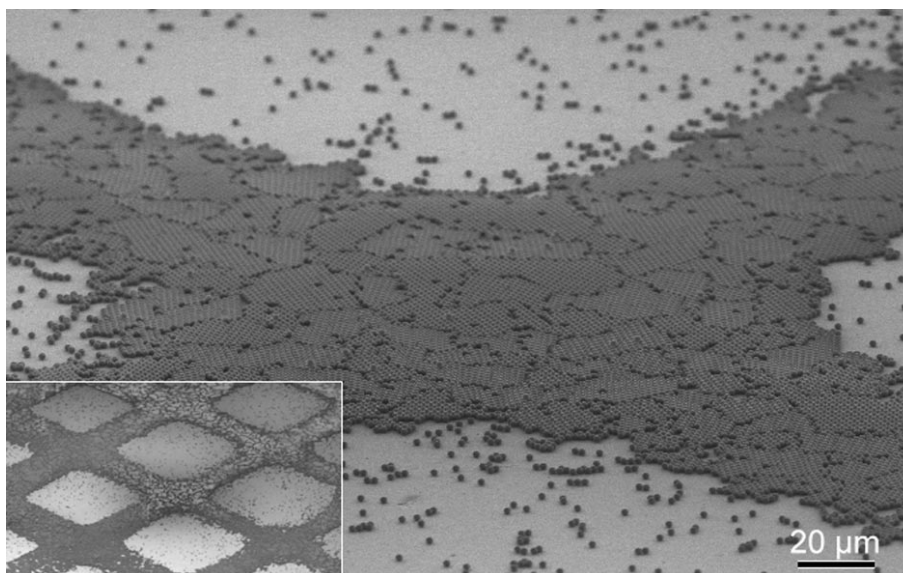


Figure 3 Scanning electron microscope image of a pattern produced by method B. Note that dense-packing of particles results in crystalline (ordered) domains that vary in size (10–20 μm). The inset ($\sim 600 \mu\text{m}$ wide) shows the overall appearance of the pattern.

Some areas shown in the inset appear white due to particle charging in the SEM. Charging in these regions may result from the higher particle density which inhibits charge transfer from the uppermost particles to the conducting substrate.

Illumination of the ITO electrode by UV light results in a small increase in the measured current through the cell when a constant voltage is applied. In our experiments, only a small area was illuminated, and the overall increase in current due to illumination was only a few per cent. As Trau *et al.*^{15,17} note, this current results from a redox reaction of water which forms hydrogen and oxygen at the electrodes. This conforms with the observation of gas-bubble formation in the cell at current densities greater than $100 \mu\text{A cm}^{-2}$. van den Meerakker *et al.*²⁴ suggest that the increase in cell current on illumination with UV light is a result of the generation—at the ITO/water interface—of hole–electron pairs, and their migration. This is consistent with statements by Bocarsly *et al.*¹⁴ that an n-type semiconductor at anodic potentials undergoes band bending at a solid–liquid interface. Electrons at the ITO surface are at a higher energy level than those in the bulk because they are surrounded by fewer metal cations¹³. When an electron–hole pair is created in this region via absorption of a photon, the promoted electron lowers its energy by moving away from the surface. Correspondingly, the new hole in the valence band moves toward the surface, where it becomes available for surface reactions. This process increases the rate of charge transfer between solution and electrode and, therefore, the current density within the cell.

As Trau *et al.*^{15,17} first noted, particles are swept into regions of higher current density. Moreover, both the electrohydrodynamic^{15,17} and the electroosmotic theories²² argue that particle movement should increase with current density and with field strength. Thus, illumination of the ITO electrode with UV light should enhance particle migration because it causes an increase in current density. Correspondingly, an increased current implies a greater potential drop across the electrolyte, so the electric field in the bulk is stronger. As illumination increases the field strength, it follows that particles in illuminated regions should be more strongly attracted to the electrode surface.

We now describe the methods we used to produce patterned assemblies of colloids. Two methods were employed (Fig. 2). Both employ a lithographic mask—consisting of regularly arranged areas ($\sim 200\text{-}\mu\text{m}$ squares) of UV-absorbing ink, with $\sim 100\text{-}\mu\text{m}$ clear areas between them—printed on a clear acetate sheet. When the mask is inserted between the ITO electrode and a light source, only selected areas of the electrode surface are illuminated by UV light.

Method A employs a suspension with a particle concentration just sufficient to produce the desired monolayer pattern. With a filter in place to cut out UV light, a d.c. voltage (1.1 V) induces deposition onto the ITO surface over a period of ~ 30 min. As the field strength is low, the particles retain a significant lateral mobility. The filter is then removed from the broad-band microscope light source, and the lithographic pattern is inserted to induce particle assembly in the illuminated regions. Over a period of ~ 2.5 h, particles in the “dark” regions are swept into the illuminated areas and the colloidal crystals grow. When the illuminated areas are filled, the strength of the electric field is increased to 3 V to irreversibly fix particles to the surface and form a permanent pattern.

In patterns fabricated by method A, most of particles are affixed to the electrode in areas illuminated by UV light. Within the illuminated regions, small crystalline domains form (generally no more than 5–10 particles across) with many gaps and imperfections. However, a substantial number of isolated particles remain in the “dark” regions of the electrode. Within the dark regions, surface coverage declines from $\sim 50\%$ to $\sim 20\%$ after 1–2 h of illumination. Continued illumination does not alter this distribution. Evidently, particle migration into illuminated regions is not strong enough to form fully dense crystals or to completely clear particles from the dark areas.

Method B employs the same apparatus with a long-wavelength UV source. Here, the initial particle concentration is just sufficient to form a monolayer over the whole surface. Particles are first

attracted to the ITO anode by applying a potential of 1 V for ~ 30 min. Then the potential is increased to 1.5 V for 10 min to assemble particles into a dense crystalline monolayer. At this field strength, the particles are not bonded to the electrode and resuspension occurs if the potential is removed. Next, the deposit is illuminated with the long-wavelength UV light source with the lithographic mask in place, and the potential increased to 3 V for 45 min. This creates a pattern of permanently bonded particles. After disassembly, the ITO electrode is rinsed with deionized water and ethanol to remove loose particles.

Particles remaining on the electrode surface following the patterning experiments were imaged via secondary electron imaging in a scanning electron microscope (SEM). Images of a pattern constructed by method B are shown in Figure 3. The intersecting strips of particles affixed to the electrode extended over most of the 1-cm^2 area of the sample cell. Although monolayer coverage prevails in the strips, there are regions where the coverage is thicker or thinner. A smattering of single particles and small clusters are present in “empty” regions (the square holes in the pattern). The polycrystalline nature of the particle layer and the specificity of the patterning procedure are clearly evident in this image.

Although our experiments employed micrometre-sized particles of polystyrene, as well as submicrometre-sized silica, and patterns with characteristic length scales of $100 \mu\text{m}$, the technique we report here should be applicable, within the constraints of colloidal particle size and mask length scales, to near-UV lithographic patterning. Because an external UV light source is used to modulate the particle assembly process, and the location and intensity of such a light beam can be externally controlled, the technique described here offers many possibilities for orchestrating electrophoretic deposition of optically tunable patterned arrays of small particles. □

Received 27 September; accepted 30 December 1999.

- Lin, S.-Y., Chow, E., Hietala, V., Villeneuve, P. R. & Joannopoulos, J. D. Experimental demonstration of guiding and bending of electromagnetic waves in a photonic crystal. *Science* **282**, 274–276 (1998).
- Painter, O. *et al.* Two-dimensional photonic band-gap defect mode laser. *Science* **284**, 1819–1821 (1999).
- Haginoya, C., Ishibashi, M. & Koike, K. Nanostructure array fabrication with a size-controllable natural lithography. *Appl. Phys. Lett.* **71**, 2934–2936 (1997).
- Gau, H., Herminghaus, S., Lenz, P. & Lipowsky, R. Liquid morphologies on structured surfaces: from microchannels to microchips. *Science* **283**, 46–49 (1999).
- Velev, O. D. & Kaler, E. W. In situ assembly of colloidal particles into miniaturized biosensors. *Langmuir* **15**, 3693–3698 (1999).
- Andres, R. P. *et al.* Research opportunities on clusters and cluster-assembled materials - a Department of Energy, Council on Materials Science panel report. *J. Mater. Res.* **4**, 704–736 (1989).
- Inanc, I. T. & Watson, G. H. Photonic band structure of fcc colloidal crystals. *Phys. Rev. Lett.* **76**, 315–318 (1996).
- Wijnhoven, J. E. G. J. & Vos, W. L. Preparation of photonic crystals made of air spheres in titania. *Science* **281**, 802–804 (1998).
- Asher, S. A., Holtz, J., Weissman, J. & Pan, G. Mesoscopically periodic photonic-crystal materials for linear and nonlinear optics and chemical sensing. *MRS Bull.* **23**, 44–50 (1998).
- Bertone, J. E., Jiang, P., Hwang, K. S., Mittleman, D. M. & Colvin, V. L. Thickness dependence of the optical properties of ordered silica-air and air-polymer photonic crystals. *Phys. Rev. Lett.* **83**, 300–303 (1999).
- Yeh, S.-R., Seul, M. & Shraiman, B. I. Assembly of ordered colloidal aggregates by electric-field-induced fluid flow. *Nature* **386**, 57–59 (1997).
- Kim, E., Xia, Y. N. & Whitesides, G. M. Two- and three dimensional crystallization of polymeric microspheres by micromolding in capillaries. *Adv. Mater.* **8**, 245–247 (1996).
- Morrison, S. R. *Electrochemistry at Semiconductor and Oxidized Metal Electrodes* (Plenum, New York, 1980).
- Bocarsly, A. B., Tachikawa, H. & Faulkner, L. R. in *Laboratory Techniques in Electroanalytical Chemistry* 2nd edn (eds Kissinger, P. T. & Heineman, W. R.) 855–898 (Marcel Dekker, New York, 1996).
- Trau, M., Saville, D. A. & Aksay, I. A. Field-induced layering of colloidal crystals. *Science* **272**, 706–709 (1996).
- Böhmer, M. In situ observation of 2-dimensional clustering during electrophoretic deposition. *Langmuir* **12**, 5747–5750 (1996).
- Trau, M., Saville, D. A. & Aksay, I. A. Assembly of colloidal crystals at electrode interfaces. *Langmuir* **13**, 6375–6381 (1997).
- Richetti, P., Prost, J. & Barois, P. Two-dimensional aggregation and crystallization of a colloidal suspension of latex spheres. *J. Phys. Lett.* **45**, 1137–1143 (1984).
- Richetti, P., Prost, J. & Barois, P. in *Physics of Complex and Supermolecular Fluids* (eds Safran, S. A. & Clark, N. A.) 387–411 (Wiley, New York, 1987).
- Giersig, M. & Mulvaney, P. Formation of ordered 2-dimensional gold colloid lattices by electrophoretic deposition. *J. Phys. Chem.* **97**, 6334–6336 (1993).

21. Giersig, M. & Mulvaney, P. Preparation of ordered colloid monolayers by electrophoretic deposition. *Langmuir* **9**, 3408–3413 (1993).
22. Solomentsev, Y., Böhrer, M. & Anderson, J. L. Particle clustering and pattern formation during electrophoretic deposition: A hydrodynamic model. *Langmuir* **13**, 6058–6068 (1997).
23. Russel, W. B., Saville, D. A. & Schowalter, W. R. *Colloidal Dispersions* (Cambridge Univ. Press, Cambridge, UK, 1989).
24. van den Meerakker, J. E. A. M., Meulenkap, E. A. & Scholten, M. (Photo)electrochemical characterization of tin-doped indium oxide. *J. Appl. Phys.* **74**, 3282–3288 (1993).
25. Benkhelifa, F., Ashrit, P. V., Bader, G., Girouard, F. E. & Truong, V.-V. Near room temperature deposited indium tin oxide films as transparent conductors and counterelectrodes in electrochromic systems. *Thin Solid Films* **232**, 83–86 (1993).
26. Murali, K. R. *et al.* Characterization of indium tin oxide films. *Surf. Coatings Technol.* **35**, 207–213 (1988).

Acknowledgements

We thank Y. Xiao and H. F. Poon for assistance in the experimental work, and A. B. Bocarsly for discussions about semiconductor properties. This work was supported by a MURI grant from the US Army Research Office, NASA (Microgravity Science and Applications Division), and a MRSEC program of the NSF.

Correspondence and requests for materials should be addressed to I.A.A. (e-mail: iaksay@princeton.edu).

Shape control of CdSe nanocrystals

Xiaogang Peng*, Liberato Manna, Weidong Yang, Juanita Wickham, Erik Scher, Andreas Kadavanich & A. P. Alivisatos

Department of Chemistry, University of California at Berkeley, and Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

Nanometre-size inorganic dots, tubes and wires exhibit a wide range of electrical and optical properties^{1,2} that depend sensitively on both size and shape^{3,4}, and are of both fundamental and technological interest. In contrast to the syntheses of zero-dimensional systems, existing preparations of one-dimensional systems often yield networks of tubes or rods which are difficult to separate^{5–12}. And, in the case of optically active II–VI and III–V semiconductors, the resulting rod diameters are too large to exhibit quantum confinement effects^{6,8–10}. Thus, except for some metal nanocrystals¹³, there are no methods of preparation that yield soluble and monodisperse particles that are quantum-confined in two of their dimensions. For semiconductors, a benchmark preparation is the growth of nearly spherical II–VI and III–V nanocrystals by injection of precursor molecules into a hot surfactant^{14,15}. Here we demonstrate that control of the growth kinetics of the II–VI semiconductor cadmium selenide can be used to vary the shapes of the resulting particles from a nearly spherical morphology to a rod-like one, with aspect ratios as large as ten to one. This method should be useful, not only for testing theories of quantum confinement, but also for obtaining particles with spectroscopic properties that could prove advantageous in biological labelling experiments^{16,17} and as chromophores in light-emitting diodes^{18,19}.

The general preparation of spherical nanocrystals involves monitoring and manipulating the kinetics of their growth. In the case of cadmium selenide, CdSe, dimethylcadmium and selenium powder are co-dissolved in a tri-alkyl phosphine (*n*-butyl or *n*-octyl), and the solution injected into hot (340–360 °C), technical grade (90% purity) trioctyl phosphine oxide (TOPO)^{14,15}. Nucleation occurs rapidly, followed by growth (280–300 °C). At the growth temperature, surfactant molecules adsorb and desorb rapidly from the nanocrystal surface, enabling the addition (as well as removal) of atoms from the crystallites, while aggregation is suppressed by the

presence of (on average) one monolayer of surfactant at the crystallite surface.

Kinetic control is then used to manipulate the average particle size and size distribution. In earlier work¹⁴, we showed that the growth process could occur in two different modes, depending upon the concentration of the monomer present: “focusing” of the size distribution, and “defocusing”. During the focusing stage, the concentration of monomer in solution is higher than the solubilities

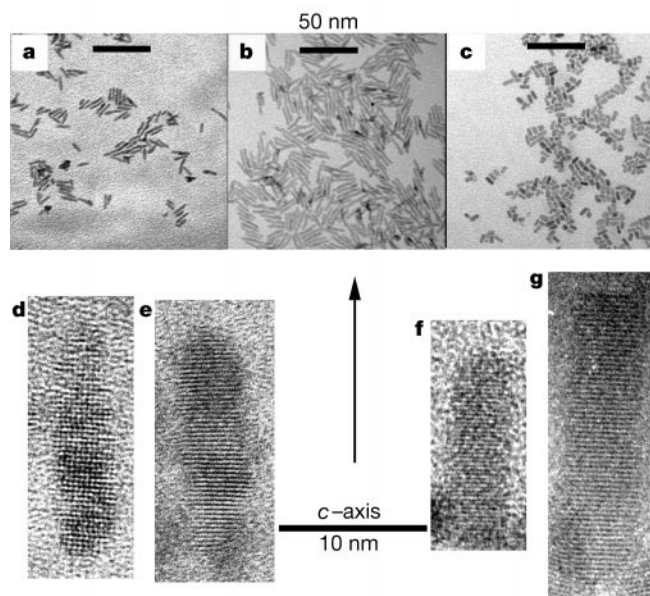


Figure 1 TEM images of different samples of quantum rods. **a–c**, Low-resolution TEM images of three quantum-rod samples with different sizes and aspect ratios. **d–g**, High-resolution TEM images of four representative quantum rods. **d** and **e** are from the sample shown in **a**; **f** and **g** are from the sample shown in **c**.

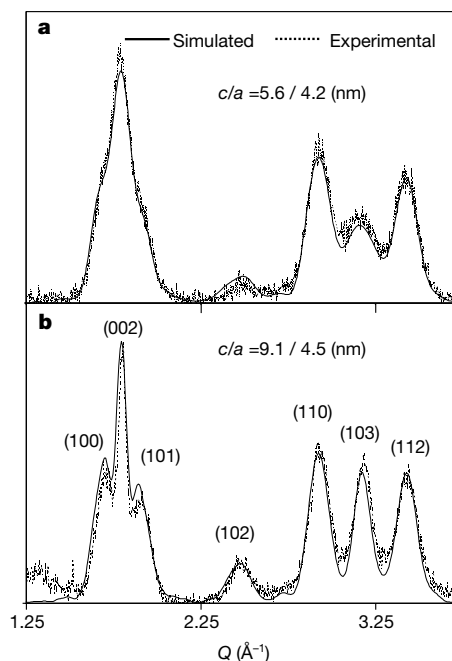


Figure 2 X-ray diffraction patterns of two CdSe rod-shaped nanocrystals. These two nanocrystals have the same short-axis dimension (within experimental error). The sizes and aspect ratios were found from the experimental patterns by the simulations shown.

* Present address: Department of Chemistry and Biochemistry, University of Arkansas, Fayetteville, Arkansas 727033, USA.

of all the particles present, a situation in which all the particles grow, regardless of size. At high monomer concentration, the smaller particles grow faster than the larger ones, and as a result, the size distribution can be focused down to one that is nearly monodisperse. If the monomer concentration drops below a critical threshold, small nanocrystals are depleted as larger ones grow and the distribution broadens, or defocuses. Thus, the preparation of nearly monodisperse spherical particles has been achieved by arresting the reaction while it is still in the focusing regime, with a large concentration of monomer still present.

Shape control of the nanocrystals can be achieved by further manipulation of the growth kinetics. This is possible because the growth of CdSe nanocrystals with the wurtzite structure ('wurtzite CdSe') is highly anisotropic when the system is kinetically over-driven by an extremely high monomer concentration. Wurtzite CdSe is intrinsically an anisotropic material, with a unique *c*-axis, and when the overall growth rate is fast, growth is generally faster along this axis. If the overall growth rate is slow, a nearly spherical, but still faceted, shape that minimizes surface area is favoured. If the growth rate is increased significantly, the result is a rod-like faceted shape where the long axis is the *c*-axis of the wurtzite crystal structure.

In order to maintain control of the growth rates in this over-driven regime, it is necessary to change the surfactants that are used. Pure TOPO is not a suitable surfactant for growing rod-like CdSe in a controllable manner. Growth in pure TOPO occurs so quickly at the high monomer concentrations desired for the growth of rods that the resulting rod-like particles are often insoluble and too big in all three dimensions. Technical grade TOPO contains additional components²⁰, which apparently act to slow the growth. Rather than rely on the fortuitous presence of impurities, addition of a molecule that coordinates more strongly than TOPO to cadmium can be used to adjust the growth rate, and hence the shape of the nanocrystals. The impurities present in technical grade TOPO which are most likely to bind relatively strongly to cadmium ions are alkyl phosphonic and phosphinic acids^{21,22}; therefore, hexyl-phosphonic acid (HPA, $C_6H_{15}PO_3$) was added to pure TOPO to 'simulate' the presence of those impurities.

In a typical synthesis, 2 ml of stock solution (Se:Cd(CH₃)₂:tributylphosphine = 1:2:38 by weight) was quickly (in $\ll 1$ s) injected into 4 g of TOPO (or TOPO + HPA) at 360 °C (or

310 °C, or 280 °C). After the injection, the temperature dropped to 300 °C (or 280 °C, or 250 °C), and aliquots were taken to monitor the reaction by ultraviolet-visible and photoluminescence spectroscopy. The reaction was stopped by removal of the heating mantle. Multiple injections were done with the same stock solution when needed. The volume of the stock solution for a secondary injection was less than 40% of the stock solution injected previously, in order to avoid the formation of new nuclei. With low concentrations of HPA in pure TOPO, 1.5% and 3% by weight, the reaction generates monodisperse spherical nanocrystals and the reaction kinetics are very similar to those reported previously¹⁴. However, higher concentrations of HPA (5%, 10% and 20%) reproducibly produce the rod-like morphologies. The best results in terms of high aspect ratios were obtained by injecting a 2-ml stock solution (Se:Cd(CH₃)₂:tributylphosphine = 1:2.6:48 by weight) into a solution of 8% HPA in TOPO at 360 °C. We will call these particles 'quantum rods', as they are still quantum-confined along two of the three axes. The aspect ratio, size and growth rate of the quantum rods can be systematically controlled by varying the reaction time, the injection and growth temperatures, and the number of injections.

We have studied the growth kinetics of the quantum rods. The average aspect ratio increases quickly just after injection, while the length of the short axis remains nearly constant (Fig. 1a, b, d and e). This stage is considered to be strongly over-driven growth, as the monomer concentration is high after the initial injection. As the

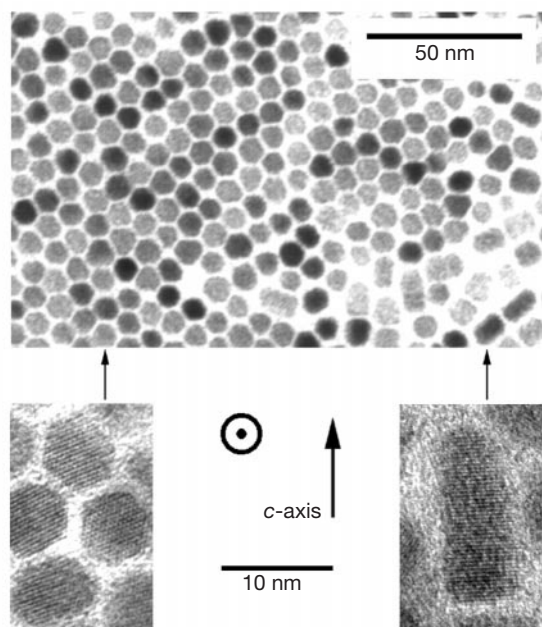


Figure 3 Three-dimensional orientation of CdSe quantum rods observed by TEM.

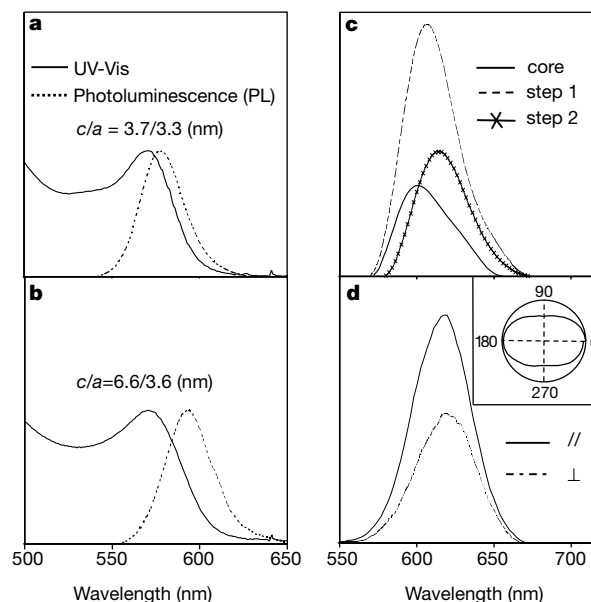


Figure 4 Optical spectra and emission polarization of CdSe quantum dots and rods.

a, Ultraviolet-visible (UV-Vis) and photoluminescence (PL) spectra of a CdSe quantum dot. **b**, UV-Vis and photoluminescence spectra of CdSe quantum rods which have short-axis lengths comparable to the diameters of the spherical dots in **a**. **c**, Photoluminescence spectra of core/shell quantum rods with different shell thickness. (Spectra were recorded on samples with the same optical density at the excitation wavelength of 520 nm.) Step-by-step shell growth was confirmed by low-resolution TEM. The optical density of all the samples shown in this figure is 0.15 ± 0.05 . **d**, Polarization resolved PL spectra of CdSe quantum rods at 4.7 K. By embedding quantum rods into poly(vinyl-butylal) (PVB) and subsequently stretching the polymer up to four times its original length, a certain degree of alignment is obtained along the stretching direction. PL spectra are taken with the detecting polarization direction parallel ($\parallel$) and perpendicular ($\perp$) to the polymer stretching direction. Inset, polar plot of the integrated PL as a function of the angle between the detecting polarization direction and the polymer stretching direction. An integrated PL polarization anisotropy ratio of up to 1.66 is observed from quantum-rod samples prepared in this way. No polarization anisotropy is observed from quantum-dot samples prepared by the same method.

monomer concentration depletes during growth, the aspect ratio gradually decreases to nearly one (ordinary quantum dots), but the short axis grows significantly. The rod morphology can be maintained (or in the case of dots, recovered) by multiple injections, which replenish the monomer, resulting in bigger quantum rods (Fig. 1c, f and g).

Transmission electron microscopy (TEM; Fig. 1) and powder X-ray diffraction (XRD; Fig. 2) confirmed the rod morphology of the resulting products. Both measurements prove that the long axis of the quantum rods is the *c*-axis of the wurtzite structure. Diffraction pattern simulations also show a much less frequent occurrence of stacking faults in the quantum rods than in spherical CdSe nanocrystals. Perpendicular to the *c*-axis, these quantum rods look like faceted hexagons under high-resolution TEM (Fig. 3). By choosing an optimal solvent/substrate combination, the *c*-axes can be aligned on a micrometre length scale, with well defined three-dimensional orientation (Fig. 3). The ability to pack and align the rods in this way will be of use in spectroscopic studies of the rods, and possibly in applications such as light-emitting diodes (LEDs) and photovoltaic cells²³.

This ability to control the shapes of semiconductor nanocrystals affords an opportunity to further test theories of quantum confinement^{24–26}, and yields samples with desirable optical characteristics from the point of view of applications. Near degeneracies in spherical dots are predicted to be lifted by elongation of the unique axis²⁷. In preliminary studies on ensembles, we have found that the splitting between absorbing and emitting states is larger in the rods than in the dots (Fig. 4a, b). This could be very helpful in applications such as LEDs where re-absorption can be a problem. Aligned nanorods in a stretched polymer at 4.7 K show polarized emission along the long axis (Fig. 4d), and this could be helpful in biological tagging experiments where the orientation of the tag needs to be determined.

The unique-axis polarized emission from the rods is in contrast to emission from round dots, for which the *c*-axis is thought to be dark²⁸. The crossover between the two regimes is of considerable theoretical interest²⁶. The quantum yield for luminescence in as-grown rods is approximately 1% at room temperature, and increases by a factor of 5 when a shell of a larger-bandgap material (CdS or ZnS) is grown on the outside of the core, as has been done with dots^{29,30}. Such quantum yields could be sufficiently high for biological labelling experiments, but are well below the maximum of 80% we have observed in dots. When thicker shells are grown, the quantum yields decrease again (Fig. 4c), possibly due to cracking or strain at the core-shell interface.

Received 6 July; accepted 17 December 1999.

1. Heath, J. M. (ed.) *Acc. Chem. Res.* **32** (Nanoscale materials special issue) (1999).
2. Alivisatos, A. P. Semiconductor clusters, nanocrystals, and quantum dots. *Science* **271**, 933–937 (1996).
3. Lieber, C. M. One-dimensional nanostructures: Chemistry, physics + applications. *Solid State Commun.* **107**, 607–616 (1998).
4. Smalley, R. E. & Yakobson, B. I. The future of the fullerenes. *Solid State Commun.* **107**, 597–606 (1998).
5. Hu, J. T., Min, O. Y., Yang, P. D. & Lieber, C. M. Controlled growth and electrical properties of heterojunctions of carbon nanotubes and silicon nanowires. *Nature* **399**, 48–51 (1999).
6. Wang, W. Z. *et al.* Synthesis and characterization of MSe (M = Zn, Cd) nanorods by a new solvothermal method. *Inorg. Chem. Commun.* **2**, 83–85 (1999).
7. Zhu, Y., Cheng, G. S. & Zhang, L. D. Preparation and formation mechanism of silicon nanorods. *J. Mater. Sci. Lett.* **17**, 1897–1898 (1998).
8. Han, W. Q., Fan, S. S., Li, Q. Q. & Hu, Y. D. Synthesis of gallium nitride nanorods through a carbon nanotube-confined reaction. *Science* **277**, 1287–1289 (1997).
9. Routkevitch, D., Bigioni, T., Moskovits, M. & Xu, J. M. Electrochemical fabrication of cds nanowire arrays in porous anodic aluminum oxide templates. *J. Phys. Chem.* **100**, 14037–14047 (1996).
10. Trentler, T. J. *et al.* Solution-liquid-solid growth of crystalline III-V semiconductors - an analogy to vapor-liquid-solid growth. *Science* **270**, 1791–1794 (1995).
11. Nishizawa, M., Menon, V. P. & Martin, C. R. Metal nanotube membranes with electrochemically switchable ion-transport selectivity. *Science* **268**, 700–702 (1995).
12. Heath, J. R. A liquid-solution-phase synthesis of crystalline silicon. *Science* **258**, 1131–1133 (1992).
13. Ahmadi, T. S., Wang, Z. L., Green, T. C., Henglein, A. & El-Sayed, M. A. Shape-controlled synthesis of colloidal platinum nanoparticles. *Science* **272**, 1924–1926 (1996).

14. Peng, X. G., Wickham, J. & Alivisatos, A. P. Kinetics of II-VI and III-V colloidal semiconductor nanocrystal growth: "Focusing" of size distributions. *J. Am. Chem. Soc.* **120**, 5343–5344 (1998).
15. Murray, C. B., Norris, D. J. & Bawendi, M. G. Synthesis and characterization of nearly monodisperse CdE (E = S, Se, Te) semiconductor nanocrystallites. *J. Am. Chem. Soc.* **115**, 8706–8715 (1993).
16. Bruchez, M., Moronne, M., Gin, P., Weiss, S. & Alivisatos, A. P. Semiconductor nanocrystals as fluorescent biological labels. *Science* **281**, 2013–2016 (1998).
17. Chan, W. C. & Nie, S. M. Quantum dot bioconjugates for ultrasensitive nonisotopic detection. *Science* **281**, 2016–2018 (1998).
18. Schlamp, M. C., Peng, X. G. & Alivisatos, A. P. Improved efficiencies in light emitting diodes made with CdSe(CdS) core/shell type nanocrystals and a semiconducting polymer. *J. Appl. Phys.* **82**, 5837–5842 (1997).
19. Mattoussi, H. *et al.* Electroluminescence from heterostructures of poly(phenylene vinylene) and inorganic CdSe nanocrystals. *J. Appl. Phys.* **83**, 7965–7974 (1998).
20. Kolosky, M. & Vialle, J. Determination of triethylphosphine oxide and its impurities by reversed-phase high performance liquid chromatography. *J. Chromatogr.* **299**, 436–444 (1984).
21. Cortina, J. L., Miralles, N., Aguilar, M. & Sastre, A. M. Distribution studies of Zn(II), Cu(II) and Cd(II) with Lewextrel resins containing di(2,4,4-trimethylpentyl) phosphinic acid (Lewatit TP807/84). *Hydrometallurgy* **40**, 195–206 (1996).
22. Kabay, N. *et al.* Removal of metal pollutants (Cd(II) and Cr(III)) from phosphoric acid solutions by chelating resins containing phosphonic or diposphonic groups. *Ind. Eng. Chem. Res.* **37**, 2541–2547 (1998).
23. Huynh, W., Peng, X. & Alivisatos, A. P. CdSe nanocrystal rods/poly(3-hexylthiophene) composite photovoltaic devices. *Adv. Mater.* **11**, 923–927 (1999).
24. Zunger, A. Electronic-structure theory of semiconductor quantum dots. *Mater. Res. Bull.* **23**, 35–42 (1998).
25. Leung, K., Pokrant, S. & Whaley, K. B. Exciton fine structure in CdSe nanoclusters. *Physical Rev. B* **57**, 12291–12301 (1998).
26. Efros, A. L. *et al.* Band-edge exciton in quantum dots of semiconductors with a degenerate valence band - dark and bright exciton states. *Phys. Rev. B* **54**, 4843–4856 (1996).
27. Nirmal, M. *et al.* Observation of the dark exciton in CdSe quantum dots. *Phys. Rev. Lett.* **75**, 3728–3731 (1995).
28. Empedocles, S. A., Neuhauser, R. & Bawendi, M. G. Three-dimensional orientation measurements of symmetric single chromophores using polarization microscopy. *Nature* **399**, 126–130 (1999).
29. Peng, X. G., Schlamp, M. C., Kadavanich, A. V. & Alivisatos, A. P. Epitaxial growth of highly luminescent CdSe/CdS core/shell nanocrystals with photostability and electronic accessibility. *J. Am. Chem. Soc.* **119**, 7019–7029 (1997).
30. Dabbousi, B. O. *et al.* (CdSe)/ZnS core-shell quantum dots: Synthesis and characterization of a size series of highly luminescent nanocrystallites. *J. Phys. Chem. B* **101**, 9463–9475 (1997).

Acknowledgements

This work was supported by the US Department of Energy and by the National Renewable Energy Laboratory.

Evidence from U–Th dating against Northern Hemisphere forcing of the penultimate deglaciation

Gideon M. Henderson*† & Niall C. Slowey‡

* Lamont-Doherty Earth Observatory of Columbia University, Route 9W, Palisades, New York 10964, USA

‡ Department of Oceanography, Texas A&M University, College Station, Texas 77843-3146, USA

Milankovitch proposed that summer insolation at mid-latitudes in the Northern Hemisphere directly causes the ice-age climate cycles¹. This would imply that times of ice-sheet collapse should correspond to peaks in Northern Hemisphere June insolation. But the penultimate deglaciation has proved controversial because June insolation peaks 127 kyr ago whereas several records of past climate suggest that change may have occurred up to 15 kyr earlier^{2–8}. There is a clear signature of the penultimate deglaciation in marine oxygen-isotope records. But dating this event, which is significantly before the ¹⁴C age range, has not been possible. Here we date the penultimate deglaciation in a record from the Bahamas using a new U–Th isochron technique. After the necessary

† Present address: Department of Earth Sciences, University of Oxford, South Parks Road, Oxford OX1 3PR, UK.

corrections for α -recoil mobility of ^{234}U and ^{230}Th and a small age correction for sediment mixing, the midpoint age for the penultimate deglaciation is determined to be 135 ± 2.5 kyr ago. This age is consistent with some coral-based sea-level estimates, but it is difficult to reconcile with June Northern Hemisphere insolation as the trigger for the ice-age cycles. Potential alternative driving mechanisms for the ice-age cycles that are consistent with such an early date for the penultimate deglaciation are either the variability of the tropical ocean-atmosphere system or changes in atmospheric CO_2 concentration controlled by a process in the Southern Hemisphere.

The timing of the rapid shifts in global climate and sea level that accompany the end of glacial periods is an important test of proposed driving mechanisms for the ice-age cycle. The widely

quoted SPECMAP timescale⁹ for Pleistocene ice-volume variations (as recorded by marine $\delta^{18}\text{O}$) uses a value of 127 kyr ago as the midpoint of the penultimate such deglacial event. This was based on (1) early U-Th dating of corals formed during the sea-level high-stand following the deglaciation, and (2) the assumption, after Milankovitch¹, that variations of Northern Hemisphere insolation at 60°N caused the changes in sea level. Recent work has suggested this age may be too late. Further dating of coral terraces has placed the peak of sea level at 128–122 kyr (ref. 2), and several studies have suggested that sea level was above the modern value as early as 135 kyr (refs 2–8). Ice-flow age models for the Vostok ice core also typically place the penultimate deglaciation at >127 kyr, although there is uncertainty about which model is most valid^{10–13}. And a record of terrestrial climate preserved in a cave deposit at Devils

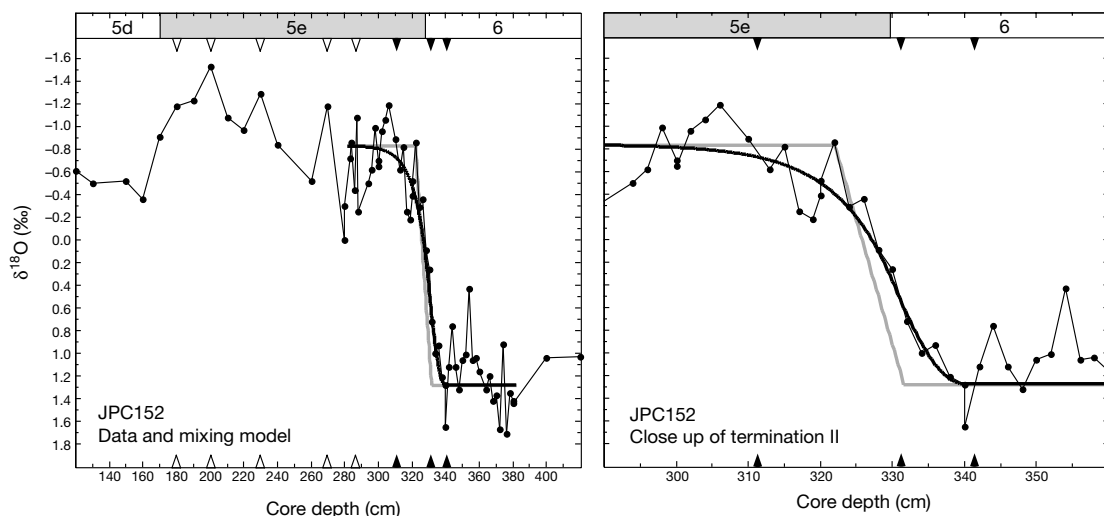


Figure 1 Oxygen-isotope stratigraphy for Bahamas core JPC152 on two different depth scales. Data points are the average of at least two replicate measurements, each on three or four *Globigerinoides sacculifer*. Shaded bars above the figures show the marine (oxygen) isotope stages (MIS). Black arrowheads mark the depths of samples dated in this study, and white arrowheads are bulk sample ages from Slowey *et al.*¹⁶. Sediment mixing increases the apparent duration and midpoint age of the deglaciation. These effects are pronounced for Bahamas slope sediments because of the rapid increase in sedimentation rate between glacial and interglacial times. The mixed layer on the slopes of the Little Bahama Bank is 8 cm thick³⁰. Solid lines in the figure are the unmixed (grey) and mixed (black) $\delta^{18}\text{O}$ curves from a simple forward model assuming efficient mixing through an 8-cm layer. Mixing is applied to both the $\delta^{18}\text{O}$ and to the age of the sediment. The mixing

model assumes a sedimentation rate of 1.5 cm kyr^{-1} in MIS 6 and 12 cm kyr^{-1} in MIS 5e (derived from the thickness of MIS 6 and MIS 5e in JPC152¹⁶). The switch in sedimentation rate is assumed to occur when full MIS 5e conditions are reached as this is approximately the timing of bank flooding. Secondary details of the deglaciation cannot be reconstructed with only three dated horizons so we assume a simple linear change in $\delta^{18}\text{O}$ from MIS 6 to MIS 5e. The duration of this change is adjusted to give a good fit to the data. Durations of between 4 and 7 kyr provide a reasonable fit to the $\delta^{18}\text{O}$ data, and also fit well with the isochron ages in this study (Fig. 3a). Adjusting parameters such as sedimentation rates in the model may alter the deglacial duration somewhat, but any reasonable input results in a midpoint of the deglaciation of 135.2 ± 2.5 kyr.

Table 1 U-Th results and 2σ errors from Bahamas core JPC152

Core-depth	^{238}U (p.p.m.)	$(^{234}\text{U}/^{238}\text{U})$	^{232}Th (p.p.b.)	$(^{230}\text{Th}/^{232}\text{Th})$ measured	f^{234}	$(^{230}\text{Th}/^{232}\text{Th})$ corrected
<i>JPC152-311cm</i>						
2.70–2.75 g cm ⁻³	2.7746 $\pm$ 0.0023	1.1628 $\pm$ 0.0033	210.8 $\pm$ 0.7	42.01 $\pm$ 0.22	1.195 $\pm$ 0.011	34.45 $\pm$ 0.42
2.75–2.80 g cm ⁻³	4.5767 $\pm$ 0.0037	1.1037 $\pm$ 0.0028	148.7 $\pm$ 0.3	80.65 $\pm$ 0.33	1.006 $\pm$ 0.007	80.07 $\pm$ 0.73
>2.80 g cm ⁻³	3.8656 $\pm$ 0.0041	1.0975 $\pm$ 0.0030	102.4 $\pm$ 0.4	91.03 $\pm$ 0.39	0.986 $\pm$ 0.008	92.45 $\pm$ 0.89
bulk sediment	4.7984 $\pm$ 0.0036	1.1012 $\pm$ 0.0027				
<i>JPC152-331cm</i>						
2.70–2.75 g cm ⁻³	0.7203 $\pm$ 0.0004	1.1345 $\pm$ 0.0034	222.9 $\pm$ 1.2	10.03 $\pm$ 0.10	1.106 $\pm$ 0.010	8.96 $\pm$ 0.13
2.75–2.80 g cm ⁻³	1.1456 $\pm$ 0.0010	1.1203 $\pm$ 0.0032	291.1 $\pm$ 2.7	11.98 $\pm$ 0.16	1.062 $\pm$ 0.009	11.20 $\pm$ 0.19
>2.80 g cm ⁻³	2.6079 $\pm$ 0.0019	1.0855 $\pm$ 0.0030	62.8 $\pm$ 0.1	97.39 $\pm$ 0.51	0.953 $\pm$ 0.007	102.79 $\pm$ 0.96
Pteropods	0.5645 $\pm$ 0.0003	1.1353 $\pm$ 0.0027	242.6 $\pm$ 1.1	7.84 $\pm$ 0.09	1.108 $\pm$ 0.008	6.99 $\pm$ 0.11
Bulk sediment	2.2356 $\pm$ 0.0014	1.1134 $\pm$ 0.0027				
<i>JPC152-341cm</i>						
2.70–2.75 g cm ⁻³	0.5037 $\pm$ 0.0002	1.1221 $\pm$ 0.0032	237.8 $\pm$ 0.9	7.45 $\pm$ 0.07	1.071 $\pm$ 0.009	6.89 $\pm$ 0.09
2.75–2.80 g cm ⁻³	1.0329 $\pm$ 0.0007	1.1178 $\pm$ 0.0034	299.7 $\pm$ 2.0	11.46 $\pm$ 0.20	1.058 $\pm$ 0.009	10.76 $\pm$ 0.22
>2.80 g cm ⁻³	2.5105 $\pm$ 0.0014	1.0900 $\pm$ 0.0035	69.1 $\pm$ 0.1	90.11 $\pm$ 0.32	0.975 $\pm$ 0.008	92.70 $\pm$ 0.88
Pteropods	0.5112 $\pm$ 0.0002	1.1247 $\pm$ 0.0029	239.3 $\pm$ 1.0	7.40 $\pm$ 0.10	1.079 $\pm$ 0.008	6.80 $\pm$ 0.12
Bulk sediment	1.1827 $\pm$ 0.0006	1.0995 $\pm$ 0.0037				

Core location: latitude 26.2267°N , longitude 77.6708°W , 577 m. f^{234} is the fraction of radiogenic ^{234}U which remains in the subsample (that is, values <1 mean that some ^{234}U is lost by recoil, while values >1 mean addition). Corrected $(^{230}\text{Th}/^{232}\text{Th})$ values are adjusted for recoil effects on ^{230}Th using f^{234} (see Methods).

Hole, Nevada, places the midpoint of the deglaciation at 142 kyr (refs 14, 15). Devils Hole is the best dated of these records, but linking the stable-isotope signal preserved there to global climate is problematic. If global climate did change significantly before 127 kyr, the penultimate deglaciation could not have been caused simply by high June Northern Hemisphere insolation, and a more complex linkage between insolation and climate must be invoked. The penultimate deglaciation is well preserved in marine $\delta^{18}\text{O}$ records but dating these records beyond the ^{14}C age range has not been possible, preventing an assessment of the timing of this important event.

Aragonite-rich sediments from the slopes of the Bahamas have recently yielded precise U–Th ages for the last interglacial¹⁶. With the bulk sediment technique used to obtain those ages, however, it was not possible to date the deglaciation preceding this interglacial. The main analytical problem to overcome in order to extend dating to this deglaciation is the separation of three sources of ^{230}Th . Radiogenic ^{230}Th , of interest for dating, must be distinguished from detrital ^{230}Th and from ^{230}Th scavenged from sea water. These last two contaminating sources of ^{230}Th also contribute ^{232}Th to the sediment, suggesting the possibility of U–Th isochron dating but, unfortunately, the two sources of contamination do not have the same $^{232}\text{Th}/^{230}\text{Th}$ ratio. Before an isochron technique can be applied, therefore, one of the sources of contamination must be accurately corrected for, or completely removed. An earlier attempt to correct isochrons for detrital ^{230}Th using aluminium measurements resulted in large corrections and uncertainty on the final age that was unacceptably high¹⁷. Here we instead quantitatively remove detrital material before analysis by a careful washing protocol. The resulting washed samples contain a mixture of radiogenic and scavenged Th which can be distinguished using an isochron.

To create the range of U/Th ratios required to construct isochrons from the washed samples, heavy-liquid separation was used to concentrate aragonite, high-Mg calcite, and low-Mg calcite into three subsamples. To assess whether these subsamples are the same age as one another, a prerequisite for isochron dating, separation was performed on a sample from the last glacial. AMS (accelerator

mass spectrometry) ^{14}C dates for the three fractions were $25,300 \pm 90$ years, $25,200 \pm 190$ years and $24,100 \pm 160$ years. The reasonable agreement between these ages indicates that using such separates for isochron dating is valid. The similarity of the ages to one another also argues against the presence of allochthonous grains in the sediment. Such allochthonous material was minimized by selecting a core from an area where no downslope transport is observed to have occurred in the latest Quaternary¹⁸. The agreement in $\delta^{18}\text{O}$ stratigraphy measured on foraminifera and fine-fraction sediment at this location provides further evidence that all sediment is directly deposited¹⁶. In order to test the isochron approach, two samples from within the ^{14}C age range were dated with the full approach outlined below. One sample was from the Holocene and the other from close to the midpoint of the most recent deglaciation. Isochron ages for these two samples are 3.3 ± 0.8 kyr and 13.0 ± 0.6 kyr ago. These dates agree with the ^{14}C chronology of the core from which samples were taken, and the age for the deglacial sample fits well with the timing of this event¹⁹. This gives us confidence that the isochron technique provides reliable ages.

To date the penultimate deglaciation, the resolution of the $\delta^{18}\text{O}$ stratigraphy for this interval was first improved (Fig. 1). Samples were then selected from the end of marine oxygen isotope stage (MIS) 6, halfway through the deglaciation, and the beginning of MIS 5e (Methods; Table 1). For each of these three isochrons, subsample ($^{234}\text{U}/^{238}\text{U}$), where parentheses indicate activity ratio, varies systematically with U/Th ratio (Fig. 2). This variation cannot be explained by the presence of detrital material, even if some has escaped the pretreatment, as subsamples with low U/Th have high ($^{234}\text{U}/^{238}\text{U}$). When compared to the ($^{234}\text{U}/^{238}\text{U}$) expected at ~ 130 kyr, subsamples with high U/Th are impoverished in ^{234}U while low-U/Th samples are enriched. Addition or removal of U from the sediment cannot readily explain this systematic variation. As a test for such addition or removal of U, however, bulk sediment ($^{234}\text{U}/^{238}\text{U}$) was also analysed for each sample. Two of the samples have ($^{234}\text{U}/^{238}\text{U}$) within error of that expected, and the third is reasonably close. Older samples from the same core also show bulk sediment ($^{234}\text{U}/^{238}\text{U}$) close to that expected¹⁶. These bulk sediment

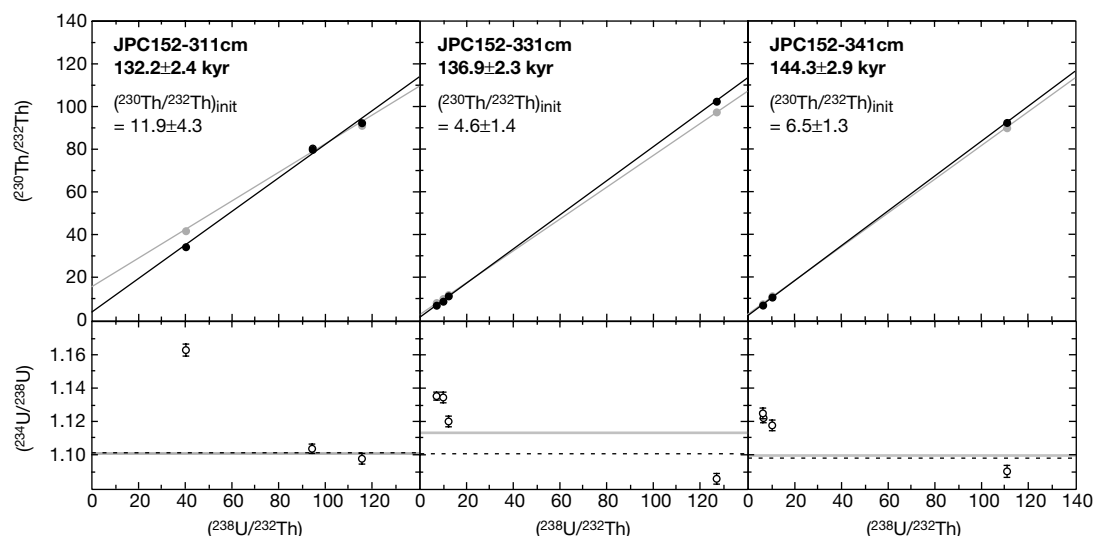


Figure 2 Isochrons and U-isotope ratios from the penultimate deglaciation. Upper panels, U–Th isochrons for three horizons in Bahamas core JPC152 (Fig. 1). Grey points and lines are measured data, black points and lines are data after correction for the effects of α -recoil (see Methods). $^{238}\text{U}/^{232}\text{Th}$ isochrons rather than $^{234}\text{U}/^{238}\text{U}$ isochrons are plotted because of the mobility of ^{234}U in these sediments. The initial thorium-isotope ratio, $(^{230}\text{Th}/^{232}\text{Th})_{\text{init}}$, calculated from the MIS 5e (311 cm) isochron, is within error of the seawater value for the Holocene¹⁶, increasing confidence that non-radiogenic Th is

derived solely from sea water. Lower panels, subsample ($^{234}\text{U}/^{238}\text{U}$) versus U/Th illustrating the effects of α -recoil. The dashed line represents the expected ($^{234}\text{U}/^{238}\text{U}$) for sediment of the age given by the isochron above. Grey lines represent the measured ($^{234}\text{U}/^{238}\text{U}$) for bulk sediment. The reasonable agreement between expected and measured ($^{234}\text{U}/^{238}\text{U}$) supports the idea that subsample deviations in ($^{234}\text{U}/^{238}\text{U}$) are due to α -recoil redistribution of ^{234}U .

data indicate that, although individual grains show altered ($^{234}\text{U}/^{238}\text{U}$), sediment at a centimetre scale has remained a closed system.

The systematic variation of ($^{234}\text{U}/^{238}\text{U}$) is well explained by internal reorganization of ^{234}U due to α -recoil. Subsamples with high U/Th values preferentially lose ^{234}U while low-U/Th subsamples gain ^{234}U . As the energy of decay of ^{234}U and ^{238}U are rather similar, ^{230}Th will be redistributed in a manner similar to ^{234}U . We are fortunate that both ^{234}U and ^{230}Th form from the decay of U which is initially incorporated with a uniform ($^{234}\text{U}/^{238}\text{U}$) ratio. This means that, regardless of the distribution of U within the grains, the recoil mobility of ^{234}U (from ^{238}U decay) tells us directly about the recoil mobility of ^{230}Th (from ^{234}U decay). The process of recoil exchange is dependent only on the rate of decay of the two U nuclides. It should also be noted that the initial decay product is an isotope of thorium in both cases (^{234}Th and ^{230}Th). So any exchange that occurs via pore waters rather than by direct grain-to-grain exchange will be chemically identical for both daughters and fast (due to the very strong tendency for thorium to adhere to particles). Deviations of the ($^{234}\text{U}/^{238}\text{U}$) ratio from that expected can therefore be used to make an entirely self-consistent correction to the ^{230}Th concentrations in the subsamples (see Methods). This correction has the effect of steepening the $^{230}\text{Th}/^{238}\text{U}$ isochrons by 15%, 7% and 4% respectively for the three isochrons (Fig. 2). The difference in the magnitude of the effect between the three isochrons is well explained by the grain size of the sediment, which is significantly finer in highstand sediment. Overall, the size of the recoil effect is larger than would be expected if U were distributed uniformly within grains of this size (63–250 μm). This observation is consistent with evidence that much of the U in these Bahamas sediments is contained in an organic phase which coats the grains²⁰.

Isochron ages corrected for the effects of α -recoil are 132.2 ± 2.4 kyr, 136.9 ± 2.3 kyr and 144.3 ± 2.9 kyr ago. Uncertainties are 2σ and are fully discussed in the Methods section. Several arguments support the validity of these ages. The internal consistency of the data argues against diagenesis, which would raise the bulk ($^{234}\text{U}/^{238}\text{U}$) value and destroy the linearity of the isochrons. The youngest isochron yields an age which fits well with the bulk-sediment U–Th age from directly above it in the same core¹⁶. This isochron requires the largest correction for recoil, so its realistic age argues strongly for the validity of the correction. Unlike corals, these samples have never been exposed to air or been in contact with meteoric water. The sediment environment can be directly investigated by measurement of U isotopes and concentrations in the pore waters of Bahamas sediment²⁰. Such analyses show that U concentrations in pore waters are more than 1,000 times lower than in the sediments analysed here. Mass balance calculations based on these porewater data indicate that, even allowing for the periodic fluid-flow which has occurred through these sediments, insignificantly small amounts of U are mobile within this sedimentary environment.

Before the isochron ages can be used to assess the timing of the penultimate deglaciation, a small correction must be made for bioturbation. Mixing increases the apparent duration and midpoint age of the deglaciation, the latter by slightly more than 1 kyr (Fig. 1). After unmixing, and allowing for uncertainty in the isochron ages, the midpoint of the penultimate deglaciation is 135.2 ± 2.5 kyr. This age is some 8 kyr older than the SPECMAP value, but is consistent with some coral data (Fig. 3). Using corals to date the end of the penultimate deglaciation directly is difficult. Relative local sea level must be corrected for the regionally variable effect of isostatic unloading²¹. Coral diagenesis and possible inaccuracy in uplift history add further complications, and mean that individual coral datum points are open to question. Corals from within ~ 30 m of peak MIS 5e sea level have, however, provided U–Th ages greater than 130 kyr in several regions^{2–8} (Fig. 3). While uncertain of the diagenetic integrity of such coral samples, Stirling *et al.*²² used an

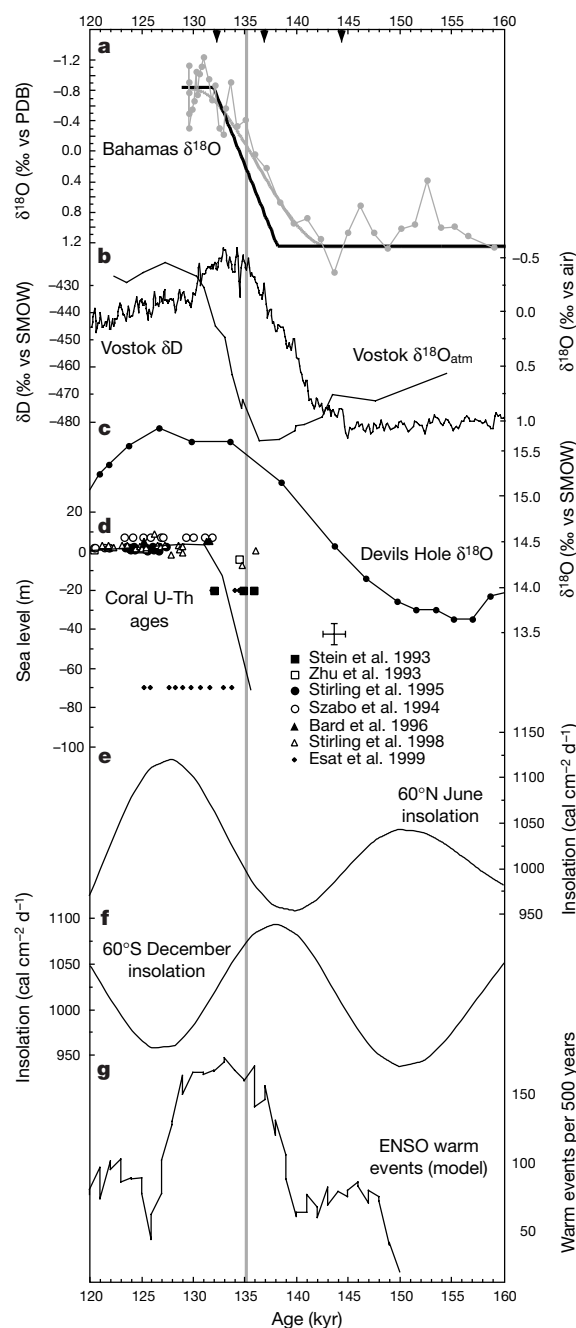


Figure 3 Timing comparison of the marine $\delta^{18}\text{O}$ record of deglaciation with data from elsewhere. **a**, $\delta^{18}\text{O}$ against age derived from U–Th isochrons at the three horizons marked by black arrowheads on the top, horizontal axis. The solid grey line is a model fit to the $\delta^{18}\text{O}$ data, and the black line is the unmixed version of this model. The midpoint of the deglaciation occurs at 135.2 ± 2.5 kyr, and is marked by the vertical grey bar. **b**, Vostok δD (ref. 11) plotted on the original Lorius timescale¹⁰ together with Vostok $\delta^{18}\text{O}$ of atmospheric O_2 , $\delta^{18}\text{O}_{\text{atm}}$ (ref. 12), plotted on a gas-age timescale consistent with the Lorius timescale²⁴. **c**, The Devils Hole $\delta^{18}\text{O}$ record¹⁵. **d**, A compilation of select coral U–Th ages versus reconstructed sea level from listed references. Only data with ($^{234}\text{U}/^{238}\text{U}$) less than or equal to 1.160 are included. The cross above the key is representative of the error on such data. No sea-level estimate was given by Bard *et al.* but these samples were from the last interglacial terrace and have therefore been placed at +5 m. Black lines represent a sea-level curve for sites far from ice-loading effects based on such coral data²². The data at -70 m have been suggested to represent a dramatic reversal in sea level after the initial rise⁷. The large spread in ages for samples from a single locality may also suggest the presence of diagenesis. **e**, **f**, Insolation curves for summer months at 60°N and 60°S . **g**, Number of warm ENSO events per 500 years in an orbitally forced model of the tropical Pacific²⁷.

isostatic model to show that, taken at face value, they required a midpoint age for the penultimate rise in sea level of ~135 kyr. The broad timing agreement between this coral-based sea-level record and that derived here from marine $\delta^{18}\text{O}$ provides further support for the validity of the isochron ages derived in this study.

The deglacial age of 135 kyr can be used to test the various age models for the Vostok ice core^{10–13}. Atmospheric $\delta^{18}\text{O}$ values ($\delta^{18}\text{O}_{\text{atm}}$), as recorded in air bubbles in Vostok¹², are a possible proxy for marine $\delta^{18}\text{O}$, but lag by ~2 kyr due to the atmospheric residence time of O_2 (refs 12, 23). Early Vostok chronologies^{10,24} place the midpoint of the penultimate deglacial shift at ~133 kyr, 2 kyr later than the midpoint of the marine $\delta^{18}\text{O}$ as derived here and therefore in good agreement. This suggests that the early Lorius timescale¹⁰ for Vostok is broadly correct, and places the midpoint of warming at Vostok (as recorded by δD) at 138–139 kyr. The 142-kyr midpoint of the well-dated Devils Hole record^{14,15} is clearly older than the midpoint of the marine $\delta^{18}\text{O}$ change.

Deglaciation centred at 135.2 ± 2.5 kyr is difficult to reconcile with an ice-age cycle triggered directly by Northern Hemisphere June insolation. A more indirect role for changes in insolation must be sought. Whereas June insolation at 60°N is below average at 135 kyr and does not peak until 127 kyr, summer insolation in the Southern Hemisphere peaks at 138 kyr (Fig. 3). The change in global atmospheric CO_2 concentration closely follows δD , and is centred at 138–139 kyr, coincident with the peak in Southern Hemisphere insolation (Fig. 3). This relationship suggests that the change in CO_2 is driven by a process in the Southern Hemisphere²³. This change in CO_2 may initiate the processes that eventually lead to the collapse of the Northern Hemisphere ice sheets. Southern Hemisphere mechanisms for the ice-age cycle are also suggested by the pattern of phasing between Southern Ocean sea surface temperature changes and $\delta^{18}\text{O}$ (ref. 25), and are consistent with possible effects of sea-ice variability²⁶. An alternative to Southern Hemisphere control of ice-sheet collapse which is also consistent with the timescale derived here, is involvement of the tropical ocean–atmosphere system. Recent modelling suggests that increasing insolation leads to a larger than average number of El Niño/Southern Oscillation (ENSO) warm events, starting at ~137 kyr ago (Fig. 3)²⁷.

Methods

Detrital material was removed from ~10-g samples of Bahamas sediment by sieving and repeatedly washing at 63–250 μm . Subsamples were then separated by centrifuging in sodium polytungstate solution at the densities shown in Table 1. Pteropod samples were hand-picked from the >250- μm fraction and thoroughly cleaned. U and Th isotopes were analysed using a VG Sector-54 thermal ionization mass spectrometer. Concentrations were measured by addition of a mixed ^{229}Th – ^{236}U spike calibrated versus the secular equilibrium standard HU-1¹⁴. U-500 standards interspersed between these analyses gave $^{235}\text{U}/^{234}\text{U} = 95.76 \pm 0.37$ (2 s.d.; $n = 14$). Blanks for the entire procedure, including heavy liquid separation, are 139 pg ^{232}Th and 760 pg ^{238}U , with $^{234}\text{U}/^{238}\text{U} \approx 1$. Decay constants used are $\lambda^{234} = 2.835 \times 10^{-6}$, $\lambda^{238} = 1.551 \times 10^{-10}$ and $\lambda^{230} = 9.195 \times 10^{-6}$.

Isochrons were corrected for the effects of α -recoil as follows. There are three unknowns: initial ($^{234}\text{U}/^{238}\text{U}$), age, and the fraction of daughter product (f) that remains in the sample after recoil. We assume that all grains started with ($^{234}\text{U}/^{238}\text{U}$) equal to the modern seawater value of 1.148. The seawater value is not thought to have changed significantly over the last 200 kyr (refs 8, 28) but the effects of small changes are investigated below. This leaves two unknowns and two measured quantities ($^{230}\text{Th}/^{238}\text{U}$ and $^{234}\text{U}/^{238}\text{U}$), so we can solve for age and f .

From the equation for decay in a radioactive series assuming that $\lambda^{238} \ll \lambda^{234}$, it can be shown that:

$$f^{234} = \frac{(^{234}\text{U}/^{238}\text{U}) - 1.148 e^{-\lambda^{234}t}}{1 - e^{-\lambda^{234}t}} \quad (1)$$

where t is the age of the sample. As the average energy of ^{238}U decay (4.184 eV; ref. 29) is not identical to ^{234}U decay (4.754 eV), we correct using:

$$\frac{1 - f^{234}}{1 - f^{230}} = \frac{\alpha^{234} M^{230}}{\alpha^{230} M^{234}} \quad (2)$$

where α is alpha decay energy released on formation of the two nuclides, and M is the mass of the recoiled nuclides. We use f^{230} calculated for each subsample in the following equation, derived from the equation of decay in a radioactive series:

$$\left(\frac{^{230}\text{Th}}{^{238}\text{U}}\right) = f^{230} \left(1 - e^{-\lambda^{230}t}\right) + \frac{f^{230} \lambda^{230}}{\lambda^{230} - f^{230} \lambda^{234}} \times 0.148 \left(e^{-f^{230} \lambda^{234}t} - e^{-\lambda^{230}t}\right) \quad (3)$$

We solve these three equations iteratively: f^{230} is initially set to 1, t is calculated for each sample using an isochron and ($^{234}\text{U}/^{238}\text{U}$)_{init} = 1.148; deviations of ($^{234}\text{U}/^{238}\text{U}$) from that expected for a sample of age t are used to calculate f^{234} for each subsample (equation (1)); f^{230} is recalculated for each subsample (equation (2)); ($^{230}\text{Th}/^{232}\text{Th}$) values are recalculated using this new f^{230} (equation (3)); t is recalculated with an isochron; and the process repeated until t and f are stationary (Table 1; Fig. 2).

Quoted errors are 2σ random error on the isochron ages, and are calculated as the quadratic sum of uncertainty due to the slope of the isochron and the recoil correction. The latter is assessed from the uncertainty in the measured ($^{234}\text{U}/^{238}\text{U}$), and is the largest contributor to the final error. Additional systematic error may be introduced if detrital material survives the pretreatment, or if past seawater had a different ($^{234}\text{U}/^{238}\text{U}$) value. If detrital material is found in all subsamples of an isochron in equal amounts, it will not bias the age. But if it occurs only in the less-dense fractions, it will cause a positive initial age. The magnitude of this effect was investigated for the 144.2-kyr isochron by assuming that 5% of the initial ^{232}Th was detrital for the three low-U/Th subsamples. This makes only a 0.25-kyr change to the age, and is therefore not significant. The survival of more than 5% detrital Th through the pretreatment is unlikely, as (1) the subsamples dissolved in weak acid and as (2) isochrons yield ($^{232}\text{Th}/^{230}\text{Th}$)_{init} values which agree with modern sea water. The effect of a 2% higher seawater ($^{234}\text{U}/^{238}\text{U}$) value at the last interglacial has also been assessed. This leads to ages for the three isochrons that are only ~0.35 kyr younger.

Received 29 April; accepted 30 December 1999.

- Milankovitch, M. in *Handbuch der Klimatologie* (eds Koppen, W. & Geiger, R.) 1–176 (Gebrüder Borntraeger, Berlin, 1930).
- Stirling, C. H., Esat, T. M., Lambeck, K. & McCulloch, M. T. Timing and duration of the Last Interglacial: Evidence for a restricted interval of widespread coral reef growth. *Earth Planet. Sci. Lett.* **160**, 745–762 (1998).
- Stein, M. et al. TIMS U-series dating and stable isotopes of the last interglacial event in Papua New Guinea. *Geochim. Cosmochim. Acta* **57**, 2541–2554 (1993).
- Zhu, Z. R. et al. High-precision U-series dating of last interglacial events by mass spectrometry: Houtman Abrolhol Islands, western Australia. *Earth Planet. Sci. Lett.* **118**, 281–293 (1993).
- Szabo, B. J., Ludwig, K. R., Muhs, D. R. & Simmons, K. R. Thorium-230 ages of corals and duration of the last interglacial sea-level high stand on Oahu, Hawaii. *Science* **266**, 93–96 (1994).
- Bard, E. et al. Pleistocene sea levels and tectonic uplift based on dating of corals from Sumba Island, Indonesia. *Geophys. Res. Lett.* **23**, 1473–1476 (1996).
- Esat, T. M., McCulloch, M. T., Chappell, J., Pillans, B. & Omura, A. Rapid fluctuations in sea level recorded at Huon Peninsula during the penultimate deglaciation. *Science* **283**, 197–201 (1999).
- Gallup, C. D., Edwards, R. L. & Johnson, R. G. The timing of high sea levels over the past 200,000 years. *Science* **263**, 796–800 (1994).
- Imbrie, J. et al. in *Milankovitch and Climate* (eds Berger, A., Imbrie, J., Hays, J., Kukla, G. & Saltzman, B.) 269–305 (Reidel, Dordrecht, 1984).
- Lorius, C. et al. A 150,000-year climatic record from Antarctic ice. *Nature* **316**, 591–596 (1985).
- Jouzel, J. et al. Extending the Vostok ice-core record of palaeoclimate to the penultimate glacial period. *Nature* **364**, 407–412 (1993).
- Sowers, T. et al. A 135,000-year Vostok-Specmap common temporal framework. *Paleoceanography* **8**, 737–766 (1993).
- Petit, J. R. et al. Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature* **399**, 429–436 (1999).
- Ludwig, K. R. et al. Mass-spectrometric ^{230}Th – ^{234}U – ^{238}U dating of the Devils Hole calcite vein. *Science* **258**, 284–287 (1992).
- Winograd, I. J., Landwehr, J. M., Ludwig, K. R., Coplen, T. B. & Riggs, A. C. Duration and structure of the past four interglaciations. *Quat. Res.* **48**, 141–154 (1997).
- Slowey, N. C., Henderson, G. M. & Curry, W. B. Direct U–Th dating of marine sediments from the two most recent interglacial periods. *Nature* **383**, 242–244 (1996).
- Henderson, G. M. & Slowey, N. C. U–Th isochron dating of the marine oxygen-isotope record. *Mineral. Mag.* **62**, 602–603 (1998).
- Burns, S. & Neumann, A. C. Pelagic sedimentation on an inactive gullied slope, Northwest Providence Channel, Bahamas. *Mar. Geol.* **77**, 277–286 (1987).
- Bard, E., Hamelin, B., Fairbanks, R. G. & Zindler, A. Calibration of the ^{14}C timescale over the past 30,000 years using mass spectrometric U–Th ages from Barbados corals. *Nature* **345**, 405–410 (1990).
- Henderson, G. M., Slowey, N. C. & Haddad, G. A. Fluid flow through carbonate platforms: Constraints from $^{234}\text{U}/^{238}\text{U}$ and Cl^- in Bahamas pore-waters. *Earth Planet. Sci. Lett.* **169**, 99–111 (1999).
- Lambeck, K. & Nakada, M. Constraints on the age and duration of the last interglacial period and on sea-level variations. *Nature* **357**, 125–128 (1992).
- Stirling, C. H., Esat, T. M., McCulloch, M. T. & Lambeck, K. High-precision U-series dating of corals from Western Australia and implications for the timing and duration of the last Interglacial. *Earth Planet. Sci. Lett.* **135**, 115–130 (1995).
- Broecker, W. S. & Henderson, G. M. The sequence of events surrounding Termination II and their implications for the cause of glacial-interglacial CO_2 change. *Paleoceanography* **13**, 352–364 (1998).
- Barnola, J. M., Pimienta, P., Raynaud, D. & Korotkevich, Y. S. CO_2 -climate relationship as deduced from the Vostok ice core: A re-examination based on new measurements and on a re-evaluation of the air dating. *Tellus B* **43**, 83–90 (1991).
- Hays, J. D., Imbrie, J. & Shackleton, N. J. Variations in the Earth's orbit: Pacemaker of the ice ages. *Science* **194**, 1121–1132 (1976).
- Kim, S. J., Crowley, T. J. & Stossel, A. Local orbital forcing of Antarctic climate change during the Last Interglacial. *Science* **280**, 728–730 (1998).
- Clement, A., Seager, R. & Cane, M. Orbital controls on ENSO and the tropical climate. *Paleoceanography* **14**, 441–456 (1999).
- Henderson, G. M., Cohen, A. S. & O'Nions, R. K. $^{234}\text{U}/^{238}\text{U}$ ratios and ^{230}Th ages for Hateruma Atoll

- corals: implications for coral diagenesis and seawater $^{234}\text{U}/^{238}\text{U}$ ratios. *Earth Planet. Sci. Lett.* **115**, 65–73 (1993).
29. Ivanovich, M. & Harmon, R. S. *Uranium-series Disequilibrium: Applications to Earth, Marine, and Environmental Sciences* (Oxford Univ. Press, 1992).
30. Henderson, G. M., Lindsay, F. & Slowey, N. C. Variation in bioturbation with water depth on marine slopes: A study on the slopes of the Little Bahamas Bank. *Mar. Geol.* **160**, 105–118 (1999).

Acknowledgements

We thank D. Schrag, R. Anderson, N. Shackleton and T. Crowley for discussion and comments, and M. Yeager for technical assistance. This work was supported by the US National Science Foundation.

Correspondence and requests for materials should be addressed to G.M.H. (e-mail: gideonh@earth.ox.ac.uk).

Stable sulphate clusters as a source of new atmospheric particles

Markku Kulmala, Liisa Pirjola & Jyrki M. Mäkelä

University of Helsinki, Department of Physics, PO Box 9, FIN-00014, University of Helsinki, Finland

The formation of new atmospheric particles with diameters of 3–10 nm has been observed at a variety of altitudes and locations. Such aerosol particles have the potential to grow into cloud condensation nuclei, thus affecting cloud formation as well as the global radiation budget. In some cases, the observed formation rates of new particles have been adequately explained by binary nucleation, involving water and sulphuric acid¹, but in certain locations—particularly those within the marine boundary layer^{1,2} and at continental sites^{1,3}—observed ambient nucleation rates exceed those predicted by the binary scheme. In these locations, ambient sulphuric acid (H_2SO_4) levels are typically lower than required for binary nucleation¹, but are sufficient for ternary nucleation⁴ (sulphuric acid–ammonia–water). Here we present results from an aerosol dynamics model with a ternary nucleation scheme which indicate that nucleation in the troposphere should be ubiquitous, and yield a reservoir of thermodynamically stable clusters 1–3 nm in size. We suggest that the growth of these clusters to a detectable size (> 3 nm particle diameter) is restricted by the availability of condensable vapour. Observations of atmospheric particle formation and growth from a continental and a coastal site support this hypothesis, indicating that a growth process including ternary nucleation is likely to be responsible for the formation of cloud condensation nuclei.

The importance of atmospheric aerosols to the global radiation balance, to cloud formation, and to alleged human health effects has motivated several recent studies on aerosol formation. Bursts of recently formed particles have been observed in several regions around the world, for example, in the free troposphere^{1,5–7}, in the marine boundary layer^{2,8–11}, at coastal sites¹², in the vicinity of evaporating clouds^{11,13}, in Arctic areas^{14,15}, in urban areas and in stack plumes^{16,17}, and in boreal forests^{3,18}. Several nucleation mechanisms have been proposed to explain this particle production, along with meteorological-related nucleation-enhancement processes such as turbulent fluctuations, waves and mixing^{19,20}.

Typically, the formation of atmospheric aerosols is attributed to binary nucleation of water and sulphuric acid^{21–23}. However, the binary theory is able to predict the nucleation rates only at some extreme conditions of low temperatures, high relative humidities, small pre-existing aerosol concentrations and at high sulphuric acid concentrations. The recently developed ternary nucleation model^{4,24} of sulphuric acid–ammonia–water (H_2SO_4 – NH_3 – H_2O)

gives significantly higher nucleation rates and thus predicts nucleation under typical tropospheric sulphuric acid (10^5 – 10^7 cm^{−3}; refs 1 and 10) and ammonia (a few p.p.t.) concentrations. Although they use different thermodynamical data (vapour pressures, surface tension) both models^{4,24} are in reasonable agreement—showing, for example, the importance of sulphuric acid molecules in the nucleation process and the effect of ammonia. Model calculations indicate that ternary nucleation happens very easily in the daytime, thus raising the question: why do we not observe new particle production as often in the atmosphere?

The answer to this is that, as a result of measurement instrumentation limitations, newly formed particles of ~ 1 nm have first to grow to detectable sizes of ~ 3 nm. Whereas ternary nucleation achieves similar particle nucleation rates to the binary scheme, with approximately 1,000 times less sulphuric acid (see Fig. 1), the corresponding condensational growth of nucleated particles will take 1,000 times longer. At such a low gas-phase concentration of H_2SO_4 , it will take several days to grow the nucleated particles up to detectable sizes. Consequently, additional vapours, other than sulphuric acid and ammonia, are required to account for particle growth.

Here we present the following hypothesis, which—in principle—enables us to explain all observed particle production bursts in the atmosphere: (1) In the atmosphere, nucleation is occurring almost everywhere, at least in the daytime. The conditions in the free troposphere (cooler temperature, fewer pre-existing aerosols) will favour nucleation. (2) Nucleation maintains a reservoir of thermodynamically stable clusters (TSCs) which are too small to be detected. (3) Under certain conditions TSCs grow to detectable sizes and further to cloud condensation nuclei.

Two theoretically predictable pathways of atmospheric nucleation are ion induced²⁵ and ternary H_2SO_4 – NH_3 – H_2O (refs 4 and 24) mechanisms. The newly formed particles, whose sizes are greater than a critical cluster (about 1 nm) but are smaller than 3 nm, are called thermodynamically stable clusters. There are two possibilities for the growth of TSC particles to detectable size (> 3 nm in diameter). First, the concentration of pre-existing aerosols is low, owing perhaps to precipitation, self-coagulation of TSCs becomes significant and the growth to 3 nm size only takes an hour (see Fig. 2b). Second, if there is a high source of available condensable vapours (such as organics, inorganic acids and ammonia) growth through condensation to detectable sizes and even to Aitken mode size occurs over a timescale of about 1–2 hours (see Fig. 3).

There are distinct reasons why the existence of TSCs in the

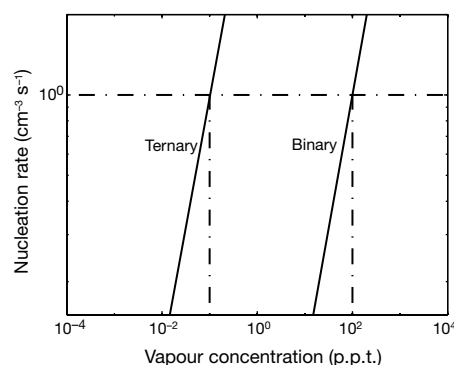


Figure 1 Nucleation rate as a function of sulphuric acid concentration for binary and ternary nucleation. The ambient sulphuric acid concentration (c_{amb}) needs to be three orders of magnitude higher in binary nucleation than in ternary nucleation in order to have the same nucleation rate (for example, $1 \text{ cm}^{-3} \text{ s}^{-1}$). The driving force in the condensational growth is proportional to $c_{\text{amb}} - c_s$, where c_s is the saturated vapour concentration at the droplet. For sulphuric acid, c_s is negligible compared to c_{amb} . Owing to the smaller vapour concentration needed for nucleation in the ternary nucleation scheme, the condensational growth of particles will be about 1,000 times slower than in the binary case.

atmosphere has not yet been observed. Almost all of the present aerosol instruments have a lower size-detection limit at a particle diameter of ~ 3 nm (refs 26 and 27), whereas TSCs are expected to be in the size range of 1–3 nm. Atmospheric ion instruments do reach the size range below 2 nm (ref. 28); however, if the TSCs are expected to be in an electrical equilibrium with the prevailing small air ions, then only an extremely small fraction (below 0.5%) can be estimated to be electrically charged²⁹. Steady concentrations in the order of 500–1,000 cm^{-3} are usually observed for small ions (sizes 1–2 nm) and bursts in the order of 1,000 cm^{-3} are seen for the

intermediate ions (sizes larger than 2 nm). This allows the number concentration of TSCs to reach a level of 100,000–200,000 cm^{-3} before making a noticeable contribution to the ion spectrum.

The hypothesis is based on theoretical calculations and model simulations. We performed more than 100 model runs using a sectional aerosol dynamic model AEROFOR³⁰ under clear-sky conditions, and the modelled results are compared with several experimental ones^{12,18}.

Figure 2 shows the particle number size distribution, total particle number concentration (N_{tot}), and the concentrations of particles

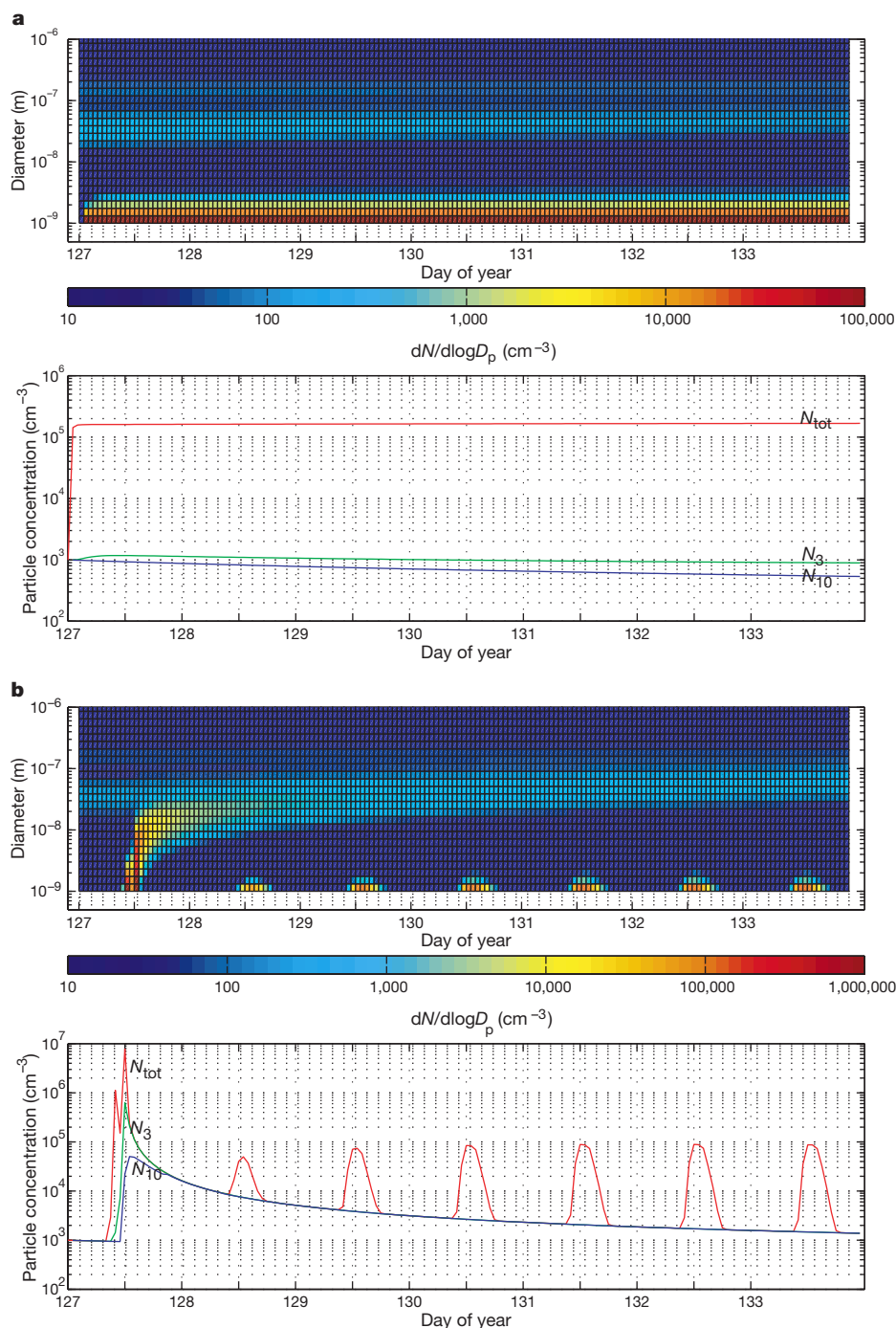


Figure 2 Simulated concentrations of particles. **a, b**, The particle number size distribution (upper panels), total particle number concentration (N_{tot}) and the concentration of particles with diameters greater than 3 nm (N_3) and 10 nm (N_{10}) (lower panels) during a seven-day simulation using constant nucleation rate (**a**), and ternary nucleation (**b**). In the upper panels the x-axis refers to time over the seven days and the y-axis to the particle diameter in each size bin; the colour bar denotes the particle number concentration in a bin (the

more red the more particles). Each size spectrum ($dN/d\log D_p$) is plotted in every 10 minutes. The initial lognormal particle size distribution was assumed to be bimodal, with parameters $N_{\text{Ait}} = 800 \text{ cm}^{-3}$, $r_{\text{Ait,dry}} = 15 \text{ nm}$ and $\sigma = 1.46$ for the Aitken mode and $N_{\text{acc}} = 200 \text{ cm}^{-3}$, $r_{\text{acc,dry}} = 65 \text{ nm}$ and $\sigma = 1.45$ for the accumulation mode. (Here r_{dry} is mean dry radius, and σ is the standard deviation in log normal particle size distribution.)

with diameters greater than 3 nm (N_3) and greater than 10 nm (N_{10}) during a seven-day period using (1) constant nucleation rate (such as ion-induced nucleation with any gases) and (2) ternary nucleation.

The total number concentration increases significantly and the steady-state value is achieved within 20 minutes (Fig. 2a). However, the amount of detectable aerosols (N_3) remains small. This means that in practice only a few of the newly formed particles (TSCs) grow and therefore, a reservoir of 1×10^5 TSCs per cm^3 exists. As very little (below 10^5 molecules cm^{-3}) condensable vapour is available in this case the insignificant growth of TSCs is due to weak self-coagulation; however, before reaching Aitken mode size, the particles are scavenged by intermode coagulation.

If nucleation is described using the ternary nucleation scheme (Fig. 2b), a reservoir of TSCs exists in the daytime. When nucleation stops a dramatic decrease in TSCs occurs. This is owing to coagulation to Aitken and accumulation mode particles. During the first day (see Fig. 2b), noticeable growth to detectable sizes owing to self coagulation of TSCs is seen, and the growth to a size of 3 nm takes only some tens of minutes. In the following days the aerosol concentration is high enough to prevent this growth of TSCs because of significantly high coagulation rates to the Aitken and accumulation modes.

Thus the steady-state concentration of TSCs (a continuous or daily value) in both cases is around $1 \times 10^5 \text{ cm}^{-3}$, and the concentrations of N_3 and N_{10} are significantly less, typically about 1,000 cm^{-3} . In Fig. 2a the constant nucleation rate is $100 \text{ cm}^{-3} \text{ s}^{-1}$; however, with a smaller value of nucleation rate the sink of TSCs is significantly greater than their source. In Fig. 2b, the daily maximum nucleation rate is $3.5 \times 10^4 \text{ cm}^{-3} \text{ s}^{-1}$ during the first day and $80\text{--}85 \text{ cm}^{-3} \text{ s}^{-1}$ during the following days, the daily maximum sulphuric acid concentration being 7.7×10^6 molecules cm^{-3} and 5×10^6 molecules cm^{-3} , respectively.

We have tested how well the hypothesis is able to explain experimental observations of atmospheric particle formation and growth under forest conditions at the Hyytiälä measurement station, Finland^{3,18} and under coastal conditions at Mace Head, Ireland¹². In both cases the agreement between theoretical values and experimental observations has been found to be good. As an example, the modelled and measured results of a typical event day at Hyytiälä are illustrated in Fig. 3. In the simulation, particle production is chosen to occur by the ternary nucleation mechanism. If the air mass with TSCs advects over a source of extra condensable vapour (for example, organic acid produced by photo-oxidation of its precursors) a rapid growth of particles to nucleation mode size and larger occurs. In this case, after 12 hours, some particles have grown even to the size of cloud condensation nuclei. It should be noted that without this extra condensable vapour the growth, through sulphuric acid condensation, of ternary nucleated TSCs is too slow, as mentioned earlier (Fig. 1). On the other hand, according to our simulations, if the extra condensable vapour is originally in the same air mass as ammonia and sulphuric acid, the nucleation will be hindered by rapid condensational growth of Aitken and accumulation mode particles. This will increase the condensational sink, reduce the sulphuric acid concentration, and prevent ternary nucleation from occurring.

The concentration of TSCs depends on the actual nucleation rate being around $10^4\text{--}10^5 \text{ cm}^{-3}$. As the size of TSCs is between the size of critical clusters (1 nm) and detectable aerosol particles (3 nm), it is difficult to observe them. In future, it would be useful to be able to observe particles of 1–3 nm in diameter.

The TSCs are able to explain observed aerosol-formation bursts in the atmosphere. The occurrence of bursts is connected to the activation of TSCs with extra condensable vapours (for example, after advection over sources or mixing). On the other hand, a small pre-existing aerosol concentration favours TSC production and

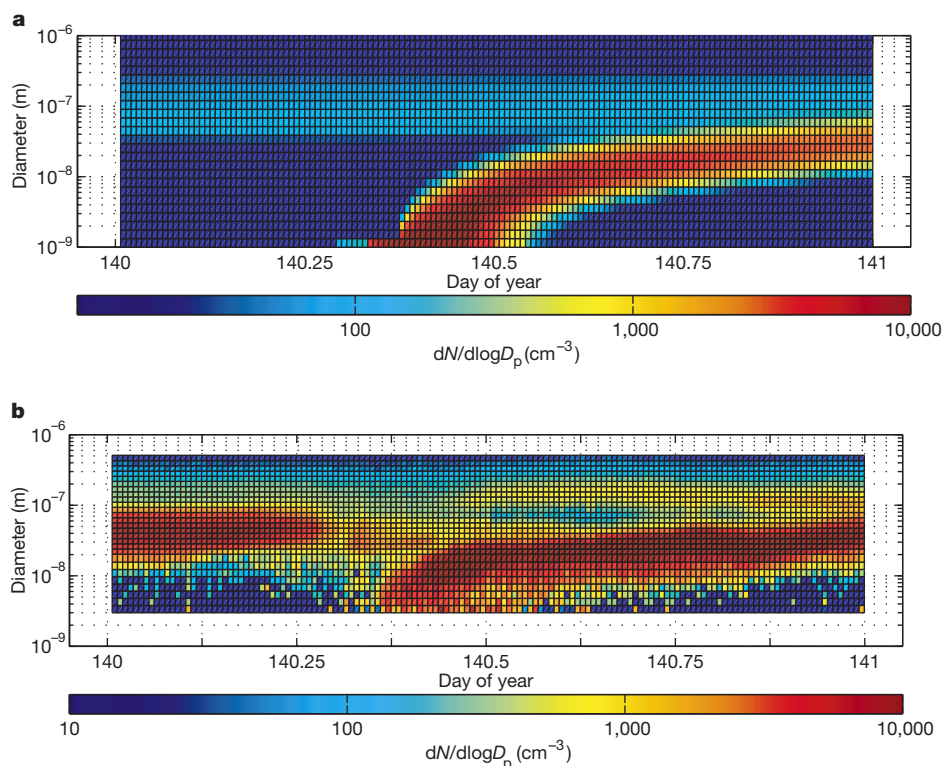


Figure 3 The modelled and experimental particle number size distributions. **a**, Modelled results; **b**, experimental values taken at Hyytiälä measurement station, Finland^{3,18} on 20 May 1998. The values used for ternary nucleation are an ammonia concentration of 15 p.p.t., and an SO_2 concentration of 100 p.p.t. OH is prescribed by semisinusoidal pattern with a maximum value of 3×10^6 and a minimum value of 3×10^3 molecules cm^{-3} . The

maximum sulphuric acid concentration is 4.4×10^6 molecules cm^{-3} . The concentration of extra condensable vapour is 3×10^7 molecules cm^{-3} , which agrees well with recent model calculations³. The simulated distributions agree well with the observed ones^{3,18}. It should be noted that before 8:00 the observed and modelled air masses are different. Axes and colour bar are as in upper panels of Fig. 2.

subsequent self-coagulation of TSCs is able to grow TSCs to a detectable size. A nucleation rate of $50\text{--}100\text{ cm}^{-3}\text{ s}^{-1}$ is needed to explain the observed number concentrations. As the ionization rate under typical tropospheric conditions is of the order of $1\text{--}5$ ion pairs $\text{cm}^{-3}\text{ s}^{-1}$ (ref. 25), we can rule out ion-induced nucleation as a probable aerosol formation route. If there exist pathways other than ternary nucleation of $\text{H}_2\text{SO}_4\text{--NH}_3\text{--H}_2\text{O}$, they should occur as easily as the ternary one. However, additional condensable vapours will still be required to activate TSCs to detectable sizes and further to cloud condensation nuclei. □

Received 5 July 1999; accepted 10 January 2000.

- Weber, R. J. *et al.* New particle formation in the remote troposphere: A comparison of observations at various sites. *Geophys. Res. Lett.* **26**, 307–310 (1999).
- Covert, D. S., Kapustin, V. N., Quinn, P. K. & Bates, T. S. New particle formation in the marine boundary layer. *J. Geophys. Res.* **97**, 20581–20587 (1992).
- Kulmala, M., Toivonen, A., Mäkelä, J. M. & Laaksonen, A. Analysis of the growth of nucleation mode particles observed in boreal forest. *Tellus* **50**, 449–462 (1998).
- Korhonen, P. *et al.* Ternary nucleation of H_2SO_4 , NH_3 , and H_2O in the atmosphere. *J. Geophys. Res.* **104**, 26349–26353 (1999).
- Clarke, A. D. Atmospheric nuclei in the remote free troposphere. *J. Atmos. Chem.* **14**, 479–488 (1992).
- Schröder, F. & Ström, J. Aircraft measurements of submicrometer aerosol particles ($> 7\text{ nm}$) in the midlatitude free troposphere and tropopause region. *Atmos. Res.* **44**, 333–356 (1997).
- Raes, F., Van Dingenen, R., Cuevas, E., Van Velthoven, P. F. J. & Prospero, J. M. Observations of aerosols in the free troposphere and marine boundary layer of the subtropical Northeast Atlantic: Discussion of processes determining their size distribution. *J. Geophys. Res.* **102**, 21315–21328 (1997).
- Hoppel, W. A., Frick, G. M., Fitzgerald, J. W. & Larson, R. E. Marine boundary layer measurements of new particle formation and the effects nonprecipitating clouds have on aerosol size distribution. *J. Geophys. Res.* **99**, 14443–14459 (1994).
- Van Dingenen, R., Raes, F. & Jensen, N. R. Evidence for anthropogenic impact on number concentration and sulphate content of cloud-processed aerosol particles over the North Atlantic. *J. Geophys. Res.* **100**, 21057–21067 (1995).
- Weber, R. J. *et al.* A study of new particle formation and growth involving biogenic trace gas species measured during ACE-1. *J. Geophys. Res.* **103**, 16385–16396 (1998).
- Clarke, A. D. *et al.* Particle production in the remote marine atmosphere: Cloud outflow and subsidence during ACE 1. *J. Geophys. Res.* **103**, 16397–16409 (1998).
- O'Dowd, C. *et al.* On the photochemical production of biogenic new particles in the coastal boundary layer. *Geophys. Res. Lett.* **26**, 1707–1710 (1999).
- Hegg, D. A., Radke, L. F. & Hobbs, P. V. Measurements of Aitken nuclei and cloud condensation nuclei in the marine atmosphere and their relation to the DMS-cloud-climate hypothesis. *J. Geophys. Res.* **96**, 18727–18733 (1991).
- Wiedensohler, A. *et al.* Occurrence of an ultrafine particle mode less than 20 nm in diameter in the marine boundary layer during Arctic summer and autumn. *Tellus B* **48**, 213–222 (1996).
- Pirjola, L., Laaksonen, A., Aalto, P. & Kulmala, M. Sulphate aerosol formation in the Arctic boundary layer. *J. Geophys. Res.* **103**, 8309–8322 (1998).
- Kerminen, V.-M. & Wexler, A. S. Post-fog nucleation of $\text{H}_2\text{SO}_4\text{--H}_2\text{O}$ particles in smog. *Atmos. Environ.* **28**, 2399–2406 (1994).
- Kerminen, V.-M. & Wexler, A. S. The occurrence of sulphuric acid-water nucleation in plumes: Urban environment. *Tellus B* **48**, 65–82 (1996).
- Mäkelä, J. M. *et al.* Observations of ultrafine aerosol particle formation and growth in boreal forest. *Geophys. Res. Lett.* **24**, 1219–1222 (1997).
- Easter, R. C. & Peters, L. K. Binary homogeneous nucleation: Temperature and relative humidity fluctuations, nonlinearity, and aspects of new particle production in the atmosphere. *J. Appl. Meteorol.* **33**, 775–784 (1994).
- Nilsson, E. D. & Kulmala, M. The potential for atmospheric mixing processes to enhance the binary nucleation rate. *J. Geophys. Res.* **103**, 1381–1389 (1998).
- Doyle, G. J. Self nucleation in the sulphuric acid-water system. *J. Chem. Phys.* **35**, 795–599 (1961).
- Raes, F., Saltelli, A. & Van Dingenen, R. Modelling formation and growth of $\text{H}_2\text{SO}_4\text{--H}_2\text{O}$ aerosols: uncertainty analysis and experimental evaluation. *J. Aerosol Sci.* **23**, 759–771 (1992).
- Kulmala, M., Laaksonen, A. & Pirjola, L. Parameterizations for sulphuric acid/water nucleation rates. *J. Geophys. Res.* **103**, 8301–8308 (1998).
- Coffman, D. J. & Hegg, D. A. A preliminary study of the effect of ammonia on particle nucleation in the marine boundary layer. *J. Geophys. Res.* **100**, 7147–7160 (1995).
- Raes, F. & Janssens, A. Ion-induced aerosol formation in a $\text{H}_2\text{O--H}_2\text{SO}_4$ system. 1. Extension of the classical theory and search for experimental evidence. *J. Aerosol Sci.* **16**, 217–227 (1985).
- Stolzenburg, M. R. & McMurry, P. H. An ultrafine aerosol condensation nucleus counter. *Aerosol Sci. Technol.* **14**, 48–65 (1991).
- Weber, R. J. *et al.* Inversion of ultrafine condensation nucleus counter pulse height distributions to obtain nonparticle (~ 3 to 10 nm) size distributions. *J. Aerosol Sci.* **29**, 601–615 (1998).
- Hörrak, U., Salm, J. & Tammet, H. Bursts of intermediate ions in atmospheric air. *J. Geophys. Res.* **103**, 13909–13915 (1998).
- Reischl, G. P., Mäkelä, J. M., Karch, R. & Nécid, J. Bipolar charging of ultrafine particles in the size range below 10 nm . *J. Aerosol Sci.* **27**, 931–949 (1996).
- Pirjola, L. Effects of the increased UV radiation and biogenic VOC emissions on ultrafine aerosol formation. *J. Aerosol Sci.* **30**, 355–367 (1999).

Acknowledgements

We thank R. J. Charlson and C. D. O'Dowd for helpful comments.

Correspondence and requests for materials should be addressed to M. K. (e-mail: markku.kulmala@helsinki.fi).

Geodetic evidence for a low slip rate in the Altyn Tagh fault system

Rebecca Bendick*, Roger Bilham*†, Jeffrey Freymueller‡, Kristine Larson§ & Guanghua Yin||

* CIREs & Department of Geological Sciences, University of Colorado, Boulder, Colorado 80309-0399, USA

† Earth Sciences, Oxford University, Parks Road, Oxford OX1 3PR, UK

‡ Geophysical Institute, PO Box 757320, University of Alaska, Fairbanks, Alaska 99775-7320, USA

§ Department of Aerospace Engineering Sciences, University of Colorado, Boulder, Colorado 80309-0429, USA

|| Xinjiang Seismic Bureau, Urumqi, Xinjiang, PRC

The collision between India and Asia has been simulated with a variety of computational models that describe or predict the motions of the main faults of east Asia. Geological slip-rate estimates of $20\text{--}30\text{ mm yr}^{-1}$ suggest that the largest of these faults, the $2,000\text{-km}$ -long Altyn Tagh fault system on the northern edge of the Tibetan plateau, absorbs as much of the Indo-Asian convergence signal as do the Himalayas^{1,2}—partly by oblique slip and partly by contraction and mountain growth^{3–5}. However, the predictions of dynamic models for Asian deformation⁶ and the lower bounds of some geological slip-rates estimates ($3\text{--}9\text{ mm yr}^{-1}$; refs 7, 8) suggest that the Altyn Tagh system is less active. Here, we report geodetic data from $89\text{--}91^\circ\text{E}$ that indicate left-lateral shear of $9 \pm 5\text{ mm yr}^{-1}$ and contraction of $3 \pm 1\text{ mm yr}^{-1}$ across the Altyn Tagh system. This result—combined with our finding that, at 90°E , Tibet contracts north–south at $9 \pm 1\text{ mm yr}^{-1}$ —supports the predictions of dynamic models of Asian deformation.

In kinematic views of the Indo-Asian plate collision, India's northward motion is absorbed by slip on a small number of active faults bounding major structural units in the collision zone^{4,9,10}. The blocks of continental crust between these major faults are usually modelled as elastic; they deform only in the sense that slip on smaller faults within them may permit slow thickening, or other relatively minor, inelastic changes of shape of each block^{9,11,12}. In contrast, dynamic models of Asian deformation (models that recognize collisional forces and inferred viscous and elastic properties of the blocks) anticipate deformation throughout the crust and upper mantle^{13,14}. Geodetic data provide present-day constraints for these models, both in the form of estimates of the translation, rotation and deformation of crustal blocks, and also in the form of independent estimates of the rate of shear strain applied to intervening faults. The geodetic surveys discussed here lie between longitudes 86° and 92°E across the Altyn Tagh fault and extend southward to India and northward to the Tien Shan. The absence of perturbations to the surface strain-field from recent earthquakes of magnitude $M > 7$ on the Altyn Tagh fault at this longitude permits us to infer the current slip-rate on the fault from the rate of shear strain centred on the fault. Levelling data record relative elevation changes of 57 buried bench-marks spaced at intervals of $4\text{--}10\text{ km}$. Benchmarks north of the Altyn Tagh fault were measured in 1957 and 1979; those south of the fault were measured in 1963 and 1979 (Fig. 1). The Global Positioning System (GPS) data include 13 control points on rock or concrete markers set at variable distances on a 300-km -long linear array across the fault.

The levelling data (Fig. 1) show a maximum uplift rate of 3 mm yr^{-1} near the northern edge of the mountains relative to points in the Tarim basin. Although a short-wavelength vertical-velocity field peaking at 1 mm yr^{-1} is evident on a $N30^\circ\text{W}$ fold belt south of the Altyn Tagh fault, no uplift occurs near the fault. We estimate a maximum vertical-velocity uncertainty in these data to be $1\text{--}2\text{ mm yr}^{-1}$ (Fig. 2) from a combination of random¹⁵ and systematic

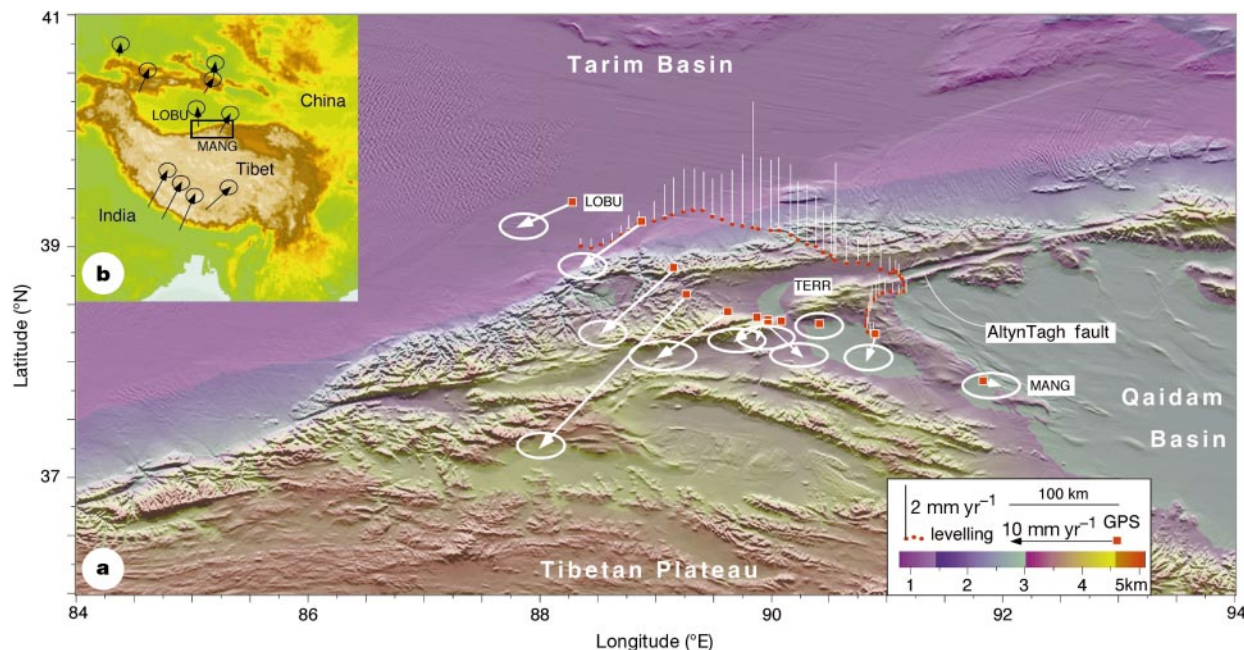


Figure 1 Deformation of the Tibetan plateau. **a**, GPS displacement vectors, and vertical velocities from spirit-level measurements, superimposed on relief map of the northern edge of the Tibetan plateau illuminated from the north. GPS velocity vectors relative to a

point (TERR) 5 km south of the Altyn Tagh fault; vertical velocities relative to the southeastern end of the levelling line (see Fig. 2b for uncertainties). **b**, GPS velocities of points between India and Tien Shan relative to Eurasia (largest vector 33 mm yr⁻¹).

errors. The absence of a significant correlation between slope and tilt-rate (Fig. 3) suggests that height-dependent errors are small. The GPS data (Table 1) were obtained as multiple 8-h sessions in 1994 and as multiple 12-h sessions in 1998 using dual frequency GPS receivers. A point at Urumqi (URUR) operated continuously during each survey and a minimum of two local stations were operated concurrently to provide short local baselines. Data from the Altyn Tagh network, plus data from fixed stations at Urumqi, Shanghai and Tsukuba were processed using the GIPSY/OASIS software¹⁶. Computational uncertainties are approximately ± 2.4 mm yr⁻¹ east and ± 1.7 mm yr⁻¹ north. With the exception of one outlier (PAXI, that we omit from consideration because its motion exceeds the displacement vectors of four neighbouring points by 2σ), the maximum fault-parallel velocity is 10 mm yr⁻¹ in a left-lateral sense relative to the mean position of points southeast of the fault, and the maximum fault-normal horizontal velocity is less than 3 mm yr⁻¹. The measured convergence vector across the Altyn Tagh fault is normal to the strike of east-southeast-trending folds south of the fault.

We assume that aseismic creep at depth is responsible for the observed surface velocity fields. If we assume further that they are generated by creep processes on single dislocations, or on relatively narrow zones of concentrated shear at depth, we may estimate the geometry and rate of slip of these surfaces using elastic dislocation theory¹⁷. The observed maxima in the vertical and horizontal surface-deformation fields are separated by more than 50 km and we model each field separately. The fault-parallel horizontal-velocity field in Fig. 2 can be emulated by slip on a vertical, buried two-dimensional fault slipping at rates of 9 ± 5 mm yr⁻¹ at locking depths of 8–36 km (1σ). The uncertainty in the rate and locking-depth is derived from the weighted misfit of the model to the data (Fig. 4).

The vertical deformation field may be modelled with the same elastic dislocation theory using a reverse fault, sub-parallel to the Altyn Tagh fault dipping southwards. Dislocation geometries that emulate the vertical-velocity field consist of slip at 5 ± 1 mm yr⁻¹ on $30 \pm 20^\circ$ SE dipping planar faults with lengths of 30–100 km

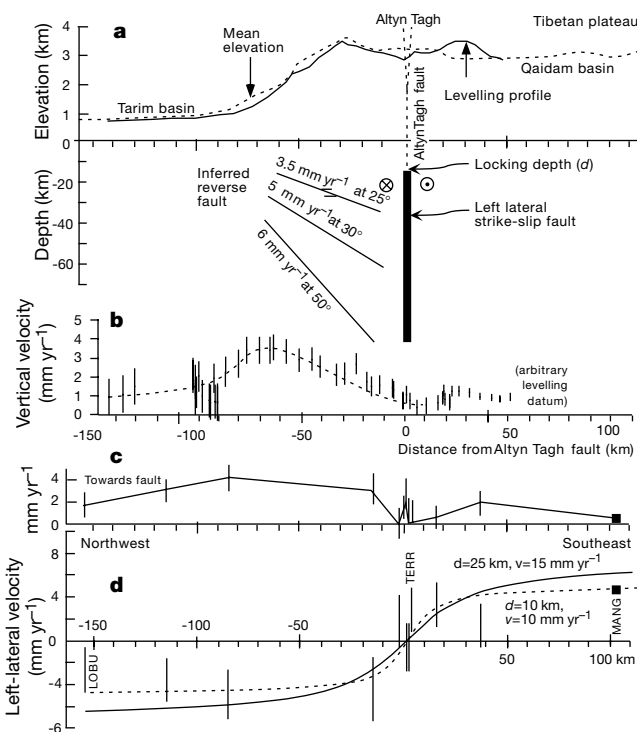


Figure 2 Sections at N10° W across the Altyn Tagh system. **a**, Topographic section normal to the Altyn Tagh fault at 90° E with observed and synthetic velocity fields projected normal to fault. **b**, Vertical velocities from levelling. **c**, Fault-normal velocities. **d**, Fault-parallel velocities from GPS geodesy. Uncertainties in the levelling data are relative to the southeastern end of the line, and uncertainties in the GPS data are shown relative to MANG. A range of inferred dislocation geometries that emulate the vertical velocity field northwest of the Altyn Tagh fault are shown with representative velocities and dips. Satisfactory (1σ) combinations of geometry and slip rate are identified in Fig. 4.

starting at a depth of 10–20 km, and with a surface projection intersecting the northern edge of the mountains. A tradeoff exists between dip and slip-rate such that the range of acceptable models result in $3 \pm 1 \text{ mm yr}^{-1}$ of surface convergence (Fig. 4), consistent with the observed fault-normal GPS contraction of $2.5 \pm 1 \text{ mm yr}^{-1}$. A common feature of these dislocations is that they must terminate down-dip, close to the vertical Altyn Tagh fault. The data are inconsistent with active dislocations that extend south of the Altyn Tagh fault, or with slip on curved surfaces.

A surprising result indicated by the 50-km separation of the maxima in the convergent and strike-slip strain fields, is that strain partitioning occurs aseismically below mid-crustal levels. In some models for partitioning, frictional conditions in the seismogenic crust control the separation of slip into dip-slip and strike-slip components^{18,19}. Our data suggest that partitioning processes are not restricted to the brittle crust, but we cannot distinguish between a mechanism where stresses driving aseismic slip are localized here by shallow seismic processes, or whether lower-crust partitioning is responsible for the stress distribution that drives shallow earthquakes. Stresses leading to strain-partitioning north of the fault are evidently similar in orientation to those responsible for folding on east-southeast axes south of the fault (Fig. 1).

The slip rate of the Altyn Tagh fault inferred from the geodetic data (Fig. 5) is 2–4 times lower than that inferred from the offset of Holocene post-glacial features³ and from ¹⁴C dating of 6 kyr BP morphological features offset by the surface fault⁵. Although the GPS data may be questioned for their low signal-to-noise ratio (4:1), we note that the slow slip rate is largely responsible for this low ratio. We believe that systematic control-point instability does not artificially lower our GPS estimate because the fault-normal GPS data are consistent with the fault-normal contraction rate inferred independently from the levelling data. Moreover, our slow slip-rate

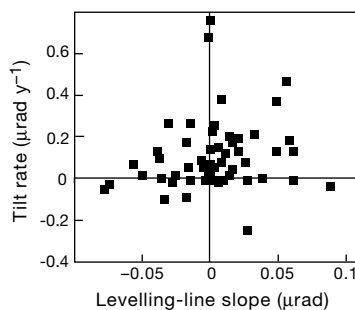


Figure 3 Plot of slope versus tilt-rate between adjacent bench marks. No clear correlation is shown, indicating a probable absence of significant systematic levelling errors.

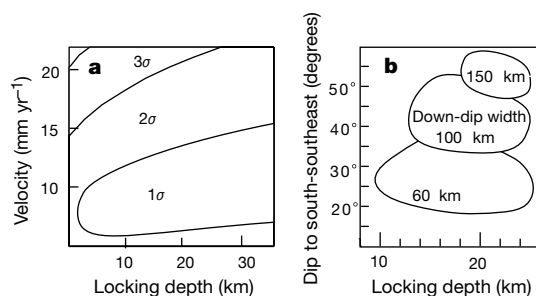


Figure 4 Solution-space contours for the dislocation models illustrated in Fig. 2. **a**, Weighted least-squares misfits to GPS observations ($1-3\sigma$). **b**, Weighted least-squares misfits to the levelling data for various locking-depths, dips, down-dip widths and slip rates.

Table 1 Coordinates and velocities of GPS points relative to MANG

Site	Latitude (°N)	Longitude (°E)	Slip rate north (mm yr ⁻¹)	Slip rate east (mm yr ⁻¹)
URUR	43.81424	87.7051	-1.6 ± 2.8	-3.3 ± 1.2
KULA	41.81643	86.19112	1.4 ± 5.0	2.1 ± 2.2
LOBU	39.44595	88.26519	-5.0 ± 3.5	-1.2 ± 1.6
MILA	39.24141	88.89869	-7.6 ± 3.2	-3.7 ± 1.5
KUMU	38.84900	89.14083	-8.4 ± 3.3	-4.4 ± 1.5
PAXI	38.61399	89.28199	-14.9 ± 3.5	-12.6 ± 1.6
NICE	38.46829	89.63004	-8.5 ± 4.4	-3.6 ± 1.9
KLSA	38.40933	89.90475	-3.2 ± 3.4	-0.9 ± 1.5
SCAN	38.40854	89.92951	2.7 ± 4.6	-3.0 ± 2.2
SCAS	38.40388	89.93227	-4.7 ± 5.3	-1.4 ± 2.9
HAPI	38.39144	89.96822	-4.9 ± 3.4	-0.9 ± 1.5
TERR	38.39113	90.08462	-1.4 ± 2.9	0.6 ± 1.2
MULI	38.37611	90.41844	-1.2 ± 3.2	-0.3 ± 1.5
HATU	38.28494	90.90676	-2.9 ± 3.1	-1.6 ± 1.4
MANG	37.88704	91.82126	0	0

on the Altyn Tagh fault is consistent with comparable slow convergence across the eastern edge of the Tibetan plateau²⁰ and with the net velocities of Altyn Tagh displacements relative to Eurasia (see inset to Fig. 1). These velocities were derived by adding velocities relative to Lhasa to the velocity of Lhasa relative to Eurasia²¹, and correcting for the rotation of Eurasia in the GPS frame of reference. In this Eurasia-fixed projection, sites southeast of the Altyn Tagh fault have a small eastward motion relative to Eurasia ($5 \pm 3 \text{ mm yr}^{-1}$), while sites northwest of the fault have a small westward motion with similar rate.

At $87-91^\circ\text{E}$, contraction between northern India and Urumqi in the northern Tien Shan amounts to $33.5 \pm 5 \text{ mm yr}^{-1}$, of which $20.3 \pm 3 \text{ mm yr}^{-1}$ occurs across the Himalaya²², $9 \pm 2 \text{ mm yr}^{-1}$ across Tibet, $2.5 \pm 1.5 \text{ mm yr}^{-1}$ across the Altyn Tagh system and $2 \pm 3 \text{ mm yr}^{-1}$ across the southeastern Tien Shan. Thus, in contrast to some interpretations of Asian kinematics in which the Altyn Tagh fault absorbs one-third of the entire Indo-Asian convergence signal, the system apparently absorbs less than 10%. The minor contraction we find in the eastern Tien Shan at 90°E , when compared with the inferred 20 mm yr^{-1} north-south contraction of the western Tien Shan²³, requires the Tarim basin to rotate clockwise relative to Eurasia at approximately 1° per Myr, a rate that must apply also to the clockwise rate of change of strike of the Altyn Tagh fault²⁴.

Although the geodetic slip rate and convergence on the Altyn Tagh system reported here is interpreted in the context of the

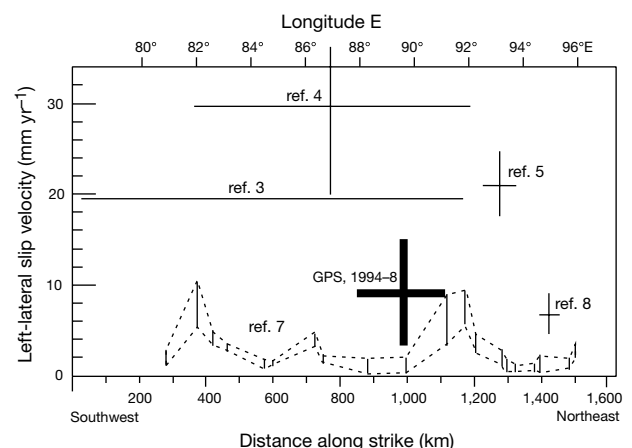


Figure 5 Dextral slip rate estimates for the Altyn Tagh fault. The geological estimates from the State Seismological Bureau⁷ indicate the range of reported values, all of which are likely to represent lower bounds. The GPS error cited is 1σ from Fig. 4a.

localized deformation implicit in kinematic reconstructions of Asian tectonics, models that incorporate Altyn Tagh slip rates faster than 15 mm yr^{-1} are inconsistent with recent GPS data. The absence of a closer agreement between current GPS rates and slip rates averaged over many thousands of years is puzzling, and if we assume each estimate to be free from error, would require a secular slowing in slip rate. The low rates reported here support the view that Tibet is currently not being extruded rapidly to the east^{13,25} and requires a larger portion of Indo-Asian convergence to be absorbed by faults elsewhere, or by internal deformation of structural units in Asia. Our findings of a low rate of shear strain on the northern margin of Tibet, and slow northward contraction and eastward extension²⁶ of the Tibetan plateau at a strain of approximately 10^{-8} yr^{-1} are similar to those predicted by dynamic models for Indo-Asian convergence. □

Received 13 August; accepted 9 December 1999.

1. Molnar, P. & Tapponnier, P. Cenozoic tectonics of Asia: Effects of a continental collision. *Science* **189**, 419–426 (1975).
2. Molnar, P. & Deng, Q. Faulting associated with large earthquakes and the average rate of deformation in central and eastern Asia. *J. Geophys. Res.* **89**, 6203–6228 (1984).
3. Peltzer, G., Tapponnier, P. & Armijo, R. Magnitude of Late Quaternary left-lateral displacements along the north edge of Tibet. *Science* **246**, 1285–1289 (1989).
4. Avouac, J.-P. & Tapponnier, P. Kinematic model of active deformation in central Asia. *Geophys. Res. Lett.* **20**, 895–898 (1993).
5. Meriaux, A. *et al.* Large-scale strain patterns, great earthquakes, and late Pleistocene slip-rate along the Altyn Tagh fault (China). *Eos (Fall Meet. Suppl.)* **79**, 400 (1998).
6. Houseman, G. & England, P. Finite strain calculations of continental collision. 1. Methods and general results for convergent zones. *J. Geophys. Res.* **91**, 3651–3663 (1986).
7. *Altyn Tagh Fault* 1–353 (Special Publication, Seismological Bureau of China, Beijing, 1992). (In Chinese.)
8. Meyer, B. *et al.* Rate of left-lateral movement along the easternmost segment of the Altyn Tagh Fault, east of 96°E China. *Geophys. J. Int.* **124**, 29–44 (1996).
9. Lamb, S. A simple method for estimating the horizontal velocity field in wide zones of active deformation—II. Examples from New Zealand, Central Asia and Chile. *Geophys. J. Int.* **119**, 313–377 (1994).
10. Peltzer, G. & Saucier, F. Present-day kinematics of Asia derived from geologic fault rates. *J. Geophys. Res.* **101**, 27943–27956 (1996).
11. Vilotte, J., Daignieres, M. & Madariaga, R. Numerical modeling of intraplate deformation: simple mechanical models of continental collision. *J. Geophys. Res.* **87**, 10709–10728 (1982).
12. Kong, X. & Bird, P. in *Tectonic Evolution of Asia* (eds Yin, A. & Harrison, T. M.) 18–34 (Cambridge Univ. Press, 1996).
13. Houseman, G. & England, P. Crustal thickening versus lateral expulsion of Tibet. *J. Geophys. Res.* **98**, 12233–12249 (1993).
14. England, P. & Molnar, P. Active deformation of Asia: from kinematics to dynamics. *Science* **278**, 647–650 (1997).
15. Jackson, M. & Bilham, R. Constraints on Himalayan deformation inferred from vertical velocity fields in Nepal and Tibet. *J. Geophys. Res.* **99**, 13897–13912 (1994).
16. Larson, K., Freymueller, J. & Phillips, S. Global plate velocities from the Global Positioning System. *J. Geophys. Res.* **102**, 9961–9982 (1997).
17. Savage, J. Displacement field for an edge dislocation in a layered half-space. *J. Geophys. Res.* **103**, 2439–2446 (1998).
18. McCaffrey, R. in *Tectonic Evolution of Southeast Asia* (eds Hall, R. & Blundell, D. J.) Vol. 106, 3–18 (Geol. Soc. Lond. Spec. Pub., 1996).
19. Jones, C. & Wesnousky, S. Variations in strength and slip rate along the San Andreas fault system. *Science* **256**, 83–86 (1992).
20. King, R. W. *et al.* Geodetic measurement of crustal motion in southwest China. *Geology* **25**, 179–182 (1997).
21. Larson, K., Bürgmann, R., Bilham, R. & Freymueller, J. Kinetics of the India-Eurasia collision zone from GPS measurements. *J. Geophys. Res.* **104**, 1111–1130 (1999).
22. Bilham, R., Larson, K., Freymueller, J. & Project Idylhim members. GPS measurements of present-day convergence across the Nepal Himalaya. *Nature* **386**, 61–64 (1997).
23. Abdurakhmatov, K. *et al.* Relatively recent construction of the Tien Shan inferred from GPS measurements of present-day crustal deformation rates. *Nature* **384**, 450–453 (1996).
24. Longjun, Y. & Liou, J. G. Two stage evolution model for the Altyn Tagh fault, China. *Geology* **27**, 227–230 (1999).
25. England, P. & Molnar, P. Right-lateral shear and rotation as the explanation for strike slip faulting in Eastern Tibet. *Nature* **344**, 140–142 (1990).
26. Bilham, R., Blume, F., Bendick, R. & Gaur, V. Geodetic constraints on the translation and deformation of India: Implications for future great Himalayan earthquakes. *Curr. Sci.* **74**, 213–229 (1998).

Acknowledgements

The investigations were funded by the National Science Foundation. G. Pelzer participated in the 1994 survey and we thank him and G. King, P. Tapponnier, P. England, and P. Molnar for discussions of the Asian collision process. R.B. received a John Simon Guggenheim Memorial Foundation fellowship while at Oxford University.

Correspondence and requests for materials should be addressed to R.B. (e-mail: bilham@stripe.colorado.edu).

Tree species impoverishment and the future flora of the Atlantic forest of northeast Brazil

José Maria Cardoso da Silva* & Marcelo Tabarelli†

* Universidade Federal de Pernambuco, Centro de Ciências Biológicas, Departamento de Zoologia, Av. Prof. Moraes Rego 1235, 50670-020, Recife, PE, Brazil

† Universidade Federal de Pernambuco, Centro de Ciências Biológicas, Departamento de Botânica, Av. Prof. Moraes Rego 1235, 50670-020, Recife, PE, Brazil

Estimates of species extinction due to human impact on tropical forests have previously been based on the relationship between species number and area¹. Here we use a different approach to estimate loss of tree species in the Atlantic forest of northeast Brazil. We evaluate the characteristics of plant species, their avian dispersers and the distribution of the forest remnants on the landscape to estimate that about 33.9% of tree species in this region will become extinct on a regional scale. Because northeast Brazil is the most threatened sector of South American Atlantic forest², our results highlight the need to change the current conservation paradigm for this region. Rather than focus on the creation of isolated reserves in any medium-to-large forest remnant, a bio regional planning approach is urgently required to rescue this unique biota from extinction.

The Atlantic forest of northeast Brazil includes all forests located north of the River São Francisco. This $35,625.92 \text{ km}^2$ region has been identified as an important area of endemism in South America^{3,4}. Its biota is influenced by the Amazonian region, making it very distinctive from other sectors of the Atlantic forest⁴. In northeast Brazil, most of the Atlantic forest has been converted into agricultural land, with only 2% of the original forest remaining^{5,6}. Forest remnants are dispersed as small patches surrounded by open fields⁶. Protected areas in this region are mostly small, isolated and badly managed². Also, hunting pressure on the fauna of these fragments is very high⁷.

Tropical forests distributed in similar landscapes to northeast Brazil are losing plant species through the disruption of key ecological processes such as pollination and seed dispersal⁸. However, no estimates of the number of threatened plant species have been made. Pollination and seed dispersal are critical because they directly affect the reproductive success of plants, and in tropical species they usually involve direct interaction with animals^{9,10}. Thus, habitat loss affects tree species through its effects on the plants themselves, on their pollinators and dispersers, or on both¹¹.

We recorded 427 tree species in the Atlantic forest of northeast Brazil. A total of 305 (71.4%) species were dispersed by vertebrates (mostly birds and mammals). Species classified as dispersed by abiotic factors represented 28.5% (122) of the total pool. We obtained data about niche regeneration for 289 species dispersed by vertebrates. Most of these (206) had fruits smaller than 15 mm. In this group, 101 (49.0%) were shade-intolerant and 93 (45.1%) were shade-tolerant species. This difference is not significant ($\chi^2 = 0.33$, $P < 0.05$). However, of the 95 vertebrate-dispersed species with fruits larger than 15 mm, there were significantly more species classified as shade-tolerant (66) than shade-intolerant (25) ($\chi^2 = 18.15$; $P < 0.002$).

A total of 78 fruit-eating birds were recorded in the region. Of these, 59 had gapes narrower than 15 mm and 19 had gapes wider than 15 mm. This difference is significant ($\chi^2 = 19.7$; $P < 0.05$). Of the species with narrow gapes, most of them (31) were edge rather than forest (28) species, but this difference is not significant ($\chi^2 = 0.15$; $P > 0.05$). A different pattern is found among species with gapes wider than 15 mm, where there were significantly more

species classified as forest (14) than edge (5) ($\chi^2 = 4.26$; $P < 0.05$). Birds can be classified qualitatively as having low, medium or high vulnerability on the basis of their responses to human disturbance¹². Narrow-gape birds inhabiting edges have mostly low vulnerability (23; 74.2%), whereas forest species have medium or high vulnerability (24; 85.7%). All wide-gape birds have medium (13) or high (6) vulnerability.

Our results suggest that, at least, 31.6% of the zoochoric trees recorded in northeast Brazil depend on wide-gape (<15 mm) birds for seed dispersal. Also there are several tree species (50; 16.3%) of the families Myrtaceae and Lauraceae that, although they produce small fruits, are dispersed preferentially by primates and large fruit-eating birds such as guans, chachalacas, toucans, aracarís and cotingas^{13–15}. In summary, about 47.9% of the zoochoric tree species in this region require dispersal by bird species that comprise only a subset (24.3%) of the regional fruit-eating avifauna. This represents 33.9% (145) of all the recorded tree species.

Of the several possible models, we recognize two critical situations for tree species and their avian dispersers in a landscape as highly fragmented as the Atlantic forest of northeast Brazil. The first consists of archipelagos of forest fragments where large fruit-eating vertebrates are now extinct; and the second of archipelagos where large fruit-eating vertebrates are present but the average distance between fragments is insuperable for them.

In the first model, local extinction of large fruit-eating birds, as well as large frugivorous mammals such as primates and agoutis, may be due to habitat loss and uncontrolled hunting. Uncontrolled hunting can lead to the local extinction of several large bird and mammal species¹⁶. The impact of heavy hunting is greater in forest patches because fragmentation may prevent hunted populations being replenished through immigration or may limit their movements across the landscape¹⁷. The Alagoas Curassow (*Mitu mitu*), the largest fruit-eating bird of the region, is extinct in the wild, possibly because of a combination of habitat loss and strong hunting pressure on its populations¹⁸. Bird species currently threatened by hunting in northeast Brazil include all large frugivores, such as guans (*Penelope*), chachalacas (*Ortalis*), toucans (*Ramphastos*) and aracarís (*Pteroglossus*). Large cotingas and trogons are sensitive to habitat disturbance and will disappear following forest fragmentation¹⁹. In a site where key vertebrate dispersers have already been extirpated, seed flow of several tree species through the landscape is very limited. One can therefore expect local extinction of tree species and their replacement by other species that produce small fruits dispersed by small, edge-living, narrow-gape birds.

In the second model, dispersal by large fruit-eating birds does not occur between patches. In this case, the distance between fragments is the main constraint on the movements of these birds between them. Most forest bird species are unable to cross more than 100–200 m of open space^{20,21} and even mammals are unable to travel large distances in an inhospitable habitat²². In this situation, one can expect tree species dispersed by wide-gape birds to be restricted to large forest fragments with high productivity that are used by large fruit-eating birds¹³. Thus, seed dispersal will be highly limited and several tree species with large fruits will become locally extinct.

If the current trend continues, the Atlantic forest on northeast Brazil will possibly be dominated by tree species dispersed by abiotic factors and tree species with small fruits. This would result in species of Melastomataceae, Rubiaceae, Myrsinaceae and Flacourtiaceae comprising most of the future flora. This 'future flora' is nowadays strongly represented in small fragments of the Atlantic forest and in early successional forests²³, habitats where large frugivores are usually absent¹⁸. These plant species are dispersed mostly by small flycatchers (Tyrannidae) and tanagers (Thraupinae), birds that are usually found in small forest fragments, forest borders and logged forests^{18,24}.

The main strategy for the conservation of Atlantic forest in northeast Brazil has been to create reserves to protect medium-to-

large forest remnants². During the 1990s, meetings have pinpointed priority areas for conservation action. In northeast Brazil, there are at least 43 recommended areas². Unfortunately, most of them contain no large fragment that, if isolated, would be able to sustain a viable population of large fruit-eating vertebrates, as the largest forest remnant in northeast Brazil is only about 2,000 ha (ref. 2). Throughout the Brazilian Atlantic forest, reserves are very small and are therefore insufficient to maintain key biological processes. Of the 239 existing reserves, over 53% (128) are smaller than 500 ha (ref. 25).

We suggest that a new paradigm for Atlantic forest conservation is urgently needed. It should not be focused on or restricted to transforming the last medium-to-large size fragments into reserves. A bioregional planning approach is required, involving the protection of landscapes composed of archipelagos of fragments, connected by corridors of original or restored vegetation and representing several thousand hectares of forest. Also, it is not enough to link fragments in local landscapes; it is essential to link landscapes at a regional level²⁶. We suggest that large fruit-eating birds should be elevated to the category of 'umbrella species'²⁷ to indicate the minimum area required to maintain the key ecological processes responsible for forest maintenance and regeneration. We believe that if area requirements for large fruit-eating birds are reached, there is a good chance of preserving other essential mutualist relationships that affect both plants and animals. □

Methods

Tree species list

A list of tree species (>2.5 cm dbh, diameter at breast height) found in the Atlantic forest of northeast Brazil (see Supplementary Information) was compiled using studies of local flora. We selected 12 studies which fulfilled the following criteria: (1) herbarium specimens are available; (2) species identification was conducted by specialists; (3) plant collection covered all the habitats inside the forest remnants. These 12 studies were conducted in sites located on the Brazilian states of Rio Grande Norte, Paraíba and Pernambuco. Tree species were first classified into two broad classes of seed-dispersal model²⁸: zoochoric species, those producing diaspores attached to a fleshy pulp, aril, or displaying features associated with vertebrate dispersal agents; and abiotic species, those adapted to wind, free fall or explosive dispersal. Zoochoric species were also grouped into two categories according to their fruit size: smaller than 15 mm, and larger than 15 mm. Dehiscent fruits were classified according to the seed length as vertebrates directly manipulate seeds. These tree species were also sub-classified by regeneration niche: shade-tolerant species, those capable of regenerating in the shaded understory of mature forest; and shade-intolerant species, those requiring high light intensity provided by treefall gaps and forest edges. Tree species were classified into these groups on the basis of (1) a reasonably good knowledge of their ecological distribution and dispersal morphology; (2) detailed published accounts of species life history traits; (3) information on fruit consumption and seed dispersal by vertebrates; and (4) checking herbarium specimens.

Bird species list

A list of fruit-eating bird species (see Supplementary Information) recorded to northeast Brazil was compiled using literature¹⁸, ornithological collections (Museu de Zoologia da Universidade de São Paulo, MZUSP, and Universidade Federal de Pernambuco, UFPE) and personal observations (J.M.C.S.). A species was classified as fruit-eating if it included a large proportion of fruits in its diet. Several flycatchers that are mostly insectivorous and supplement their diets with fruit were not included. Granivorous species, such as doves (*Leptotila*) and quails (*Odontophorus*) were excluded as they are poor seed dispersers and seeds are generally destroyed in the large, muscular gizzard during digestion². Species were classified according to the width of bill at gape into two groups, smaller than 15 mm and larger than 15 mm, because gape size limits the size of fruit that can be swallowed²⁴. Gape measurements were obtained directly from specimens of the ornithological collection of the Universidade Federal de Pernambuco, Recife, and Museu Paraense Emílio Goeldi, Belém. Habitat preferences for each species was determined from the literature¹⁸ and field data gathered by J.M.C.S. in several sites along the ranges of these species. Birds were grouped into two categories according to habitat preferences: edge species, those restricted to borders, and large gaps within the forests; and forest species, those restricted to the interior and canopy of forest. Bird species vulnerability was assessed using the classification of all South American birds according to their sensitivity to human disturbance¹².

Received 2 November 1999; accepted 21 January 2000.

- Brooks, T. & Balmford, A. Atlantic forest extinctions. *Nature* **380**, 115 (1996).
- Dias, I. S., Câmara, I. G. & Lino, C. F. *Workshop Mata Atlântica: Problemas, Diretrizes e Estratégias de Conservação* (Fundação SOS Mata Atlântica, São Paulo, 1990).
- Müller, P. *The Dispersal Centres of Terrestrial Vertebrates in the Neotropical Realm* (Junk, The Hague, 1973).
- Prance, G. T. in *Biological Diversification in the Tropics* (ed. Prance, G. T.) 137–158 (Columbia Univ. Press, New York, 1982).
- Viana, V. M., Tabanez, A. J. & Batista, J. L. in *Tropical Forest Remnants: Ecology, Management, and*

- Conservation of Fragmented Communities* (eds Laurence W. F., Laurence, R. O. & Bierregard, J.) 351–365 (Univ. Chicago Press, 1997).
6. Ranta, P., Blom, T., Niemela, J., Joensuu, E. & Siitonen, M. The fragmented Atlantic rain forest of Brazil: size, shape and distribution of forest fragments. *Biodiversity Cons.* **7**, 385–403 (1998).
 7. Almeida, R. T., Pimentel, D. S. & Silva, E. M. S. The Red-handed Howling monkey in the state of Pernambuco, north-east Brazil. *Neotropical Primates* **3**, 174–175 (1995).
 8. Corlett, R. T. & Turner, I. M. in *Tropical Forest Remnants: Ecology, Management, and Conservation of Fragmented Communities* (eds Laurence, W. F., Laurence, R. O. & Bierregard, J.) 333–346 (Chicago Univ., 1997).
 9. Wunderlee, J. M. The role of animal seed dispersal in accelerating native forest regeneration on degraded tropical lands. *Forest Ecol. Managmt.* **9**, 223–235 (1997).
 10. Murcia, C. in *Forest Patches in Tropical Landscapes* (eds Schellas, J. & Greenberg, R.) 19–36 (Island, Washington DC, 1996).
 11. Laurence, W. F. in *Tropical Forest Remnants: Ecology, Management, and Conservation of Fragmented Communities* (eds Laurence, W. F., Laurence, R. O. & Bierregard, J.) 275–280 (Chicago Univ. Press, 1997).
 12. Stotz, D. F., Fitzpatrick, J. W., Parker, T. A. & Moskovits, D. K. *Neotropical Birds: Ecology and Conservation* (Chicago Univ. Press, 1996).
 13. Guidon, F. C. in *Forest Patches in Tropical Landscapes* (eds Schellas, J. & Greenberg, R.) 163–186 (Island, Washington DC, 1996).
 14. Galetti, M., Martuscelli, P., Olmos, F. & Aleixo, A. Ecology and conservation of the jacutinga *Pipile jacutinga* in the Atlantic forest of Brazil. *Biol. Cons.* **82**, 31–39 (1997).
 15. Moraes, P. L. R. Espécies utilizadas na alimentação no mono-carvoeiro (*Brachyteles arachnoides* E. Geoffroy, 1806) no Parque Estadual Carlos Botelho. *Revista do Instituto Florestal* **4**, 1206–1208 (1992).
 16. Redford, K. H. The empty forest. *BioScience* **42**, 412–422 (1992).
 17. Robinson, J. G. in *Forest Patches in Tropical Landscapes* (eds Schellas, J. & Greenberg, R.) 111–130 (Island, Washington DC, 1996).
 18. Sick, H. *Ornitologia Brasileira: uma Introdução* (Univ. Brasília, 1985).
 19. Willis, E. O. The composition of avian communities in remanent woodlots in southern Brazil. *Papéis Avulsos de Zoologia* **33**, 1–25 (1979).
 20. Bierregaard, R. O. & Dale, V. H. in *Forest Patches in Tropical Landscapes* (eds Schellas, J. & Greenberg, R.) 187–204 (Island, Washington DC, 1996).
 21. Silva, J. M. C., Uhl, C. & Murray, G. Plant succession, landscape management, and the ecology of frugivorous birds in abandoned pastures. *Cons. Biol.* **10**, 491–503 (1996).
 22. Offerman, H. L., Dale, V. H., Pearson, S. M., Bierregaard, R. O. & O'Neill, R. V. Effects of forest fragmentation on neotropical fauna. *Environ. Rev.* **3**, 191–211 (1995).
 23. Tabarelli, M., Mantovani, W. & Peres, C. A. Effects of habitat fragmentation on plant guild structure in the montane forest of southeastern Brazil. *Biol. Cons.* **91**, 119–127 (1999).
 24. Moermond, T. C. & Denslow, J. S. Neotropical avian frugivores: patterns of behavior, morphology, and nutrition, with consequences for fruit selection. *Ornithol. Mon.* **36**, 865–897 (1985).
 25. Lima, A. R. & Capobianco, J. P. R. Mata Atlântica: avanços legais e institucionais para sua conservação. *Documentos do ISA* **4**, 1–118 (1997).
 26. Soulé, M. E. & Terborgh, J. in *Continental Conservation: Scientific Foundations of Regional Reserve Network* (eds Soulé, M. E. & Terborgh, J.) 1–17 (Island, Washington DC, 1999).
 27. Caro, T. M. & O'Doherty, G. On the use of surrogate species in conservation biology. *Cons. Biol.* **13**, 805–814.
 28. Pijl, L. Van der *Principles of Dispersal of Higher Plants* (Springer, Berlin, 1982).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank the World Wildlife Fund-Brazil and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for financial support. The text benefited from comments by V. H. Dale and E. O. Wilson.

Correspondence and request for materials should be addressed to J.M.C.S. (e-mail: jmcslva@npd.ufpe.br) or M.T. (e-mail: mtrelli@npd.ufpe.br).

Egg investment is influenced by male attractiveness in the mallard

Emma J. A. Cunningham*† & Andrew F. Russell*†

* Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK

Why females prefer to copulate with particular males is a contentious issue. Attention is currently focused on whether females choose males on the basis of their genetic quality, in order to produce more viable offspring¹. Support for this hypothesis in birds has come from studies showing that preferred males tend to

father offspring of better condition or with increased survivorship^{2–8}. Before attributing greater offspring viability to a male's heritable genetic quality, however, it is important to discount effects arising from confounding sources, including maternal effects. This has generally been addressed by comparing offspring viability from two different breeding attempts by the same female: one when offspring are sired by a preferred male, and one when offspring are sired by a less preferred male. However, here we show that individual female mallard (*Anas platyrhynchos*) lay larger eggs after copulating with preferred males and smaller eggs after copulating with less preferred males. As a result, females produced offspring of better body condition when paired with preferred males. After controlling for these differences in maternal investment, we found no effect of paternity on offspring condition. This shows that differences between half-sibs cannot always be attributed to paternal or maternal genetic effects.

Life-history theory predicts that females should alter their investment in a particular breeding attempt according to the likelihood of its success⁹. If preferred males provide any type of benefit to females, for example, better resources for breeding, individual females should alter their investment according to male attractiveness. This was first demonstrated by showing that females may increase their level of parental care when paired with preferred males^{10,11} (but see ref. 12). However, females may also alter their investment much earlier in reproduction; in some species, females lay more eggs when paired with preferred males than when paired with less preferred males^{13,14}. A problem arises if females instead alter their primary reproductive effort in ways that influence offspring quality rather than offspring quantity, for example, by laying larger eggs or increasing the nutritional content of eggs when paired with more attractive males. These effects would then have to be discounted before attributing all differences in offspring condition to viability genes inherited from the father. A recent study has shown that female zebra finches (*Taeniopygia guttata*), for example, alter their testosterone investment in eggs according to male attractiveness¹⁵. Testosterone is thought to influence early competitive ability in chicks¹⁶.

Female mallard show strong preferences for particular males, yet rear the precocial young on their own. Female preferences are clear from pairing displays performed throughout the autumn^{17–19}. Once pair bonds are formed in resident populations, they are maintained between years^{19,20}. Male mallard ($n = 20$) were ranked for attractiveness (see Methods). To ensure ranking was reliable, males were ranked twice with two separate groups of 20 female mallard and by a different observer on each occasion; male rank correlated significantly between the two groups of females (Spearman's rank correlation (r_s) = 0.61, $n = 20$, $P = 0.008$). Once male rank was established, we examined paternal and maternal effects on offspring traits by collecting three clutches of eggs from each of 16 females; an infertile control clutch, a clutch sired by a high-ranking male and a clutch sired by a low-ranking male. The two fertile clutches ($n = 32$

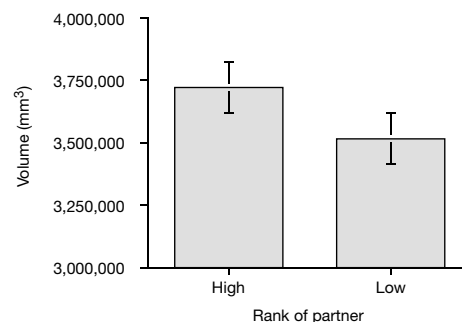


Figure 1 Mean egg volume (mm³) in high rank pairing and low rank pairing treatments (paired t -test, $t = 2.75$, $n = 16$ females, $P = 0.02$).

† Present address: Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK.

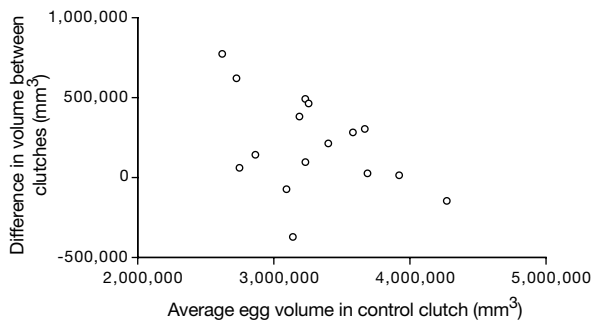


Figure 2 Females that laid smaller eggs in their initial control clutch were significantly more likely to lay eggs of different sizes when assigned to males of different attractiveness ($r_p = -0.435$, $n = 16$ females, $P < 0.05$).

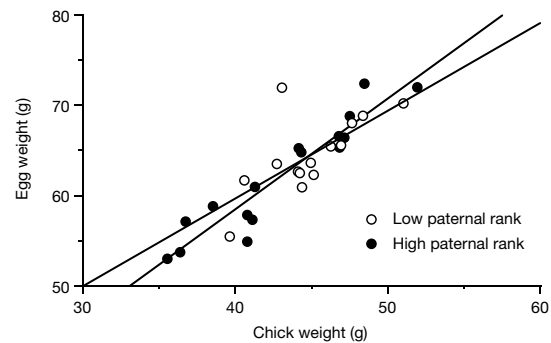


Figure 3 Relationship between mean egg weight and mean chick weight for each female ($F = 64.68$, d.f. = 14, $P < 0.001$) in high paternal rank and low paternal rank groups (analysis of covariance, $F_{\text{slopes}} = 2.62$, d.f. = 14, $P < 0.118$; $F_{\text{elevations}} = 0.16$, d.f. = 14, $P = 0.58$).

clutches, mean clutch size = 11.54, range 6–16) were incubated artificially under controlled conditions and chicks ($n = 200$) were monitored from hatching, through growth and development, until they themselves paired the following year.

We found that individual females laid significantly larger eggs when paired to more attractive males (Fig. 1). Egg size is a critical trait influencing fitness in birds^{21,22}; in the mallard, this is because young from larger eggs are better able to survive during the few days immediately after hatching²³. The differences in egg investment appeared to stem primarily from females that laid small eggs in their control clutch; females that initially laid larger eggs did not adjust egg size according to male rank (Fig. 2). However, the difference in average egg volume between clutches was unrelated to female condition (Pearson's correlation (r_p) = 0.021, $n = 16$, $P > 0.1$). There was no difference in the number of eggs laid by females for high- and low-ranking males (paired t -test, $t = 1.00$, degrees of freedom (d.f.) = 14, $P = 0.33$, mean number of eggs ($\pm$ s.d.): high rank treatment, 12.6 ± 2.8 ; low rank treatment, 11.6 ± 2.7), suggesting that females were differing in their total investment rather than trading-off egg size and clutch size. This difference in egg investment resulted in females producing significantly heavier chicks when they were assigned to preferred males than when assigned to less preferred males (paired t -test, $t = 2.39$, d.f. = 14, $P = 0.03$; mean weight ($\pm$ s.d.): high rank treatment, 45.0 ± 3.0 g; low rank treatment, 43.0 ± 4.8 g), though chicks did not differ in their skeletal size (paired t -test, $t = 0.28$, d.f. = 14, $P = 0.78$, mean tarsal length ($\pm$ s.d.): high rank treatment, 49.6 ± 2.2 mm, low rank treatment, 49.5 ± 1.9 mm). This could not be explained by chicks that were fathered by preferred males being more likely to be male; there was no difference in the sex ratio of clutches fathered by high-

and low-ranking males (paired t -test, $t = 0.07$, $n = 16$ females, $P = 0.95$, mean proportion of males to females ($\pm$ s.d.): high rank treatment = 0.47 ± 0.2 , low rank treatment = 0.47 ± 0.2). When we controlled for the difference in egg investment, there was no effect of male rank on chick weight (Fig. 3).

This is the first study to our knowledge to show that female investment, in terms of egg volume, may vary according to male attractiveness, even though males appear to provide no direct material benefits to females. It has been stated that differences in viability traits between half-sibs that share the same mother but that are sired by males of different attractiveness can be attributed to paternal genetic quality^{2,4,8}. The results of this study show that this may not be the case.

The question then is why do some females lay eggs of different sizes when assigned to preferred and less preferred males? Females may alter their egg investment in response to male attractiveness because attractive males are more likely to sire successful offspring. This could be adaptive whether the benefits of mating with preferred males are direct material benefits or indirect genetic benefits unrelated to viability, such as attractiveness genes. It is unclear why this should not be a beneficial strategy to all females, but it may be that females that initially laid smaller eggs were inherently of poorer quality and had the most to gain from increasing egg size with preferred males. Alternatively, females that initially laid larger eggs may not have been able to increase their egg size any further.

Therefore, we first examined the effect of paternal attractiveness on other measures of viability and male offspring attractiveness. We found no effect of egg size (all regressions, $P > 0.1$) or paternal rank on any other offspring traits that we recorded (Table 1). Instead, the

Table 1 Paired comparisons of traits between half-sibs sharing the same mother but fathered by preferred and less preferred males and hatching from first and second clutches.

Trait	Effect of male rank*				Effect of hatch date†			
	Mean ($\pm$ s.d.)	t	P		Mean ($\pm$ s.d.)	t	P	
Hatching success	49.5 $\pm$ 0.2 (high-rank male) 48.5 $\pm$ 0.2 (low-rank male)	0.12	0.91		44 $\pm$ 25 (1st clutch) 53 $\pm$ 22 (2nd clutch)	1.34	0.20	
Survival	82 $\pm$ 18 (high-rank male) 86 $\pm$ 17 (low-rank male)	0.41	0.69		80 $\pm$ 22 (1st clutch) 87 $\pm$ 12 (2nd Clutch)	0.39	0.71	
Attractiveness	3.8 $\pm$ 1.7 (high-rank male) 4.0 $\pm$ 1.8 (low-rank male)	0.29	0.79		4.5 $\pm$ 1.2 (1st clutch) 3.3 $\pm$ 1.8 (2nd clutch)	3.21	0.03	
Growth rate	F_{slopes} 0.64	P 0.44	$F_{\text{elevations}}$ 0.52	P 0.49	F_{slopes} 0.90	P 0.37	$F_{\text{elevations}}$ 5.00	P 0.04
Moult development	0.09	0.76	0.29	0.59	0.28	0.61	8.76	0.01

* Hatch date balanced across treatments.

† Male rank balanced across treatments.

‡ Calculated from mean values for each clutch from each female.

single most important factor determining both viability traits and male attractiveness was hatch date (first versus second clutches) (Table 1). The onset of breeding is synchronous in mallard, but up to 96% of first clutches can be lost through depredation and females can lay up to four new clutches within a season¹⁹. Attractiveness in the mallard therefore appears to be an environmentally determined trait (see also ref. 24). Hatch date is primarily governed by nest depredation, which is mainly a chance event. Any genetic benefits to be gained from copulating with preferred males are therefore unlikely to become correlated with male attractiveness through future generations, including any ability by males to persuade females to lay larger eggs.

We then examined whether direct costs and benefits of pairing with preferred males may be important. Male mallard provide no paternal care and the benefits of pairing with particular males are not immediately apparent. However, there are at least two possibilities that could account for our findings. First, females prefer early hatched males. Dominance in wildfowl is often a consequence of pairing success²⁵. Early hatched chicks are the first males to pair and are therefore likely to be established as the most dominant males. Pair males defend transient feeding areas around females during the breeding season. Combined with their better body condition arising from earlier hatching, and the fact that early pairing increases their dominance status, such males are likely to be the most effective in defending a feeding area. This possibility is unlikely to explain our results directly as birds in this study were fed *ad libitum* and kept separately in pairs, but it may be one reason why females benefit from laying larger eggs for preferred males under natural conditions.

Second, the amount a female can invest in eggs may simply be a direct consequence of being paired to a particular male. There is a tendency for less preferred males to attempt more copulations with their partner than preferred males ($r_s = 0.68$, $n = 8$, $P = 0.07$), the majority of which are resisted by females. Unwanted copulation attempts in other species can influence female fecundity²⁶ and female longevity²⁷.

The results of this experimental study clearly show that females may invest differentially in eggs depending on the attractiveness of their partner. As a consequence, maternal investment can have an important influence on offspring quality. This study shows the fundamental importance of considering all maternal effects prior to attributing differences in offspring viability to paternal genetic effects in the future. □

Methods

Establishing experimental stock

Eggs were collected from wild-caught captive stock mallard that had been fed *ad libitum* from three months before egg laying to minimize any differences in female condition. Eggs were incubated artificially under standardized conditions. On hatching, chicks were reared in isolation from their parents and again fed *ad libitum* over the experimental period to minimize any differences in condition.

Ranking of males

The following spring, one experimental group of 20 males and two groups of 20 females were established. The first group of females was housed together with the males and allowed to pair freely. Over the pairing period, males were ranked for attractiveness to females by recording the number of females that directed pairing displays toward them. This is a reliable measure of female choice rather than the outcome of male–male competition^{17–19}. Once female preferences had been established, females were removed and housed individually until egg laying commenced. A second group of naive females were introduced into the experimental pen containing the previously ranked males. Female preferences were again recorded for the same males, by a second observer who had no previous knowledge of the males' initial ranking, to determine whether different females consistently prefer the same males.

Effects of male rank on offspring viability and female investment

Females were housed individually for about two weeks until they started laying. A control clutch of infertile eggs was collected to establish the average egg size for each female without the influence of male quality affecting female investment. Eggs were collected daily, measured and their volumes calculated using species-specific calculations²⁸.

The eight highest ranking males and the eight lowest ranking males were selected. Females were assigned one of these males for their next clutch; half of the females received a male of high rank and half a male of low rank. Pairs were housed together until females had finished laying an entire clutch. Eggs were collected each day and replaced with a chicken egg to induce females to lay a normal clutch. Once females had completed laying, males were removed. The chicken eggs were removed 7–10 days later to induce females to lay another clutch. A second male was then introduced to the female pen and the pair were left together throughout the female's second clutch. When re-laying occurred, females had been separated from their first male for at least 17 days, the longest period a female mallard has been recorded to store sperm^{19,29}, so that all offspring in the second clutch could be fathered only by the second male. For the second clutch, females previously paired with high-ranking males were paired to low-ranking males and *vice versa*, and the same procedure was followed for collection and incubation of the second clutch. Hence, two clutches were obtained from each female, one fertilized by a high-ranking male, the other by a low-ranking male. Seasonal effects were equally distributed over the two treatments. All birds were fed *ad libitum* during the experiment on a mix of wheat grain and high protein crumb with calcium and mineral supplements.

Eggs were incubated artificially under standardized conditions. Immediately on hatching, chicks were ringed with individual combinations and their morphometrics were recorded. All chicks were measured and weighed at weekly intervals until reaching adult size. Male moult development was scored weekly and males were assessed for attractiveness the following spring in the same manner as their fathers.

Received 11 October; accepted 23 December 1999.

- Andersson, M. *Sexual Selection* (Princeton Univ. Press, Princeton, NJ, 1994).
- Hasselquist, D., Bensch, S. & von Schantz, T. Correlation between male song repertoire, extra-pair paternity and offspring survival in the great reed warbler. *Nature* **381**, 229–232 (1996).
- Kempnaers, B. *et al.* Extra-pair paternity results from female preference for high-quality males in the blue tit. *Nature* **357**, 494–496 (1992).
- Kempnaers, B., Verheyen, G. R. & Dhondt, A. Extra-pair paternity in the blue tit *Parus caeruleus*: female choice, male characteristics, and offspring quality. *Behav. Ecol.* **8**, 481–492 (1997).
- Møller, A. P. Male ornament size as a reliable cue to enhanced offspring viability in the barn swallow. *Proc. Natl Acad. Sci. USA* **91**, 6929–6932 (1994).
- Norris, K. Heritable variation in a plumage indicator of viability in male great tits *Parus major*. *Nature* **362**, 537–539 (1993).
- Petrie, M. Improved growth and survival of offspring of peacocks with more elaborate trains. *Nature* **371**, 598–599 (1994).
- Sheldon, B. C., Merila, J., Qvarnstrom, A., Gustafsson, L. & Ellegren, H. Paternal genetic contribution to offspring condition predicted by size of a male secondary sexual character. *Proc. Roy. Soc. Lond. B* **264**, 297–302 (1997).
- Williams, G. C. Natural selection, the cost of reproduction, and a refinement of Lack's principle. *Am. Nat.* **100**, 687–690 (1966).
- Burley, N. The differential allocation hypothesis: an experimental test. *Am. Nat.* **132**, 611–628 (1988).
- de Lope, F. & Møller, A. P. Female reproductive effort depends on the degree of ornamentation of their mates. *Evolution* **47**, 1152–1160 (1993).
- Witte, K. The differential allocation hypothesis: does the experimental evidence support it? *Evolution* **49**, 1289–1290 (1995).
- Petrie, M. & Williams, A. Peahens lay more eggs for peacocks with larger trains. *Proc. Roy. Soc. Lond. B* **251**, 127–131 (1993).
- Schwabl, H., Mock, D. W. & Gieg, J. A. A hormonal mechanism for parental favouritism. *Science* **386**, 231 (1997).
- Gil, D., Graves, J., Hazon, N. & Wells, A. Male attractiveness and differential testosterone investment in zebra finch eggs. *Science* **286**, 126–128 (1999).
- Balzer, A. L. & Williams, T. D. Do female zebra finches vary primary reproductive effort in relation to male attractiveness? *Behaviour* **135**, 297–309 (1998).
- Lorenz, K. Comparative studies on the behaviour of the Anatinae. *Avicult. Mag.* **57**, 157–182 (1951).
- Omland, K. E. Female mallard mating preferences for multiple male ornaments. I. Natural variation. *Behav. Ecol. Sociobiol.* **39**, 353–360 (1996).
- Cunningham, E. J. A. *Forced Copulation and Sperm Competition in the Mallard (Anas platyrhynchos)*. Thesis, Univ. Sheffield, UK (1997).
- Mjelstad, H. & Sætersdal, M. Reforming of resident Mallard pairs *Anas platyrhynchos*, rule rather than exception? *Wildfowl* **41**, 150–151 (1990).
- Williams, T. D. Intraspecific variation in egg size and egg composition in birds: effects on offspring fitness. *Biol. Rev.* **68**, 35–59 (1994).
- Bolton, M. Determinants of chick survival in the lesser black-backed gull: relative contributions of egg size and parental quality. *J. Anim. Ecol.* **60**, 949–960 (1991).
- Rhymer, J. M. The effect of egg size variability on thermoregulation of Mallard *Anas platyrhynchos* offspring and its implications for survival. *Oecologia* **75**, 20–24 (1988).
- Griffith, S. C., Owens, I. P. F. & Burke, T. Environmental determination of a sexually selected trait. *Nature* **400**, 358–360 (1999).
- Sorenson, L. G. & Derrickson, S. R. Sexual selection in the northern pintail (*Anas acuta*): the importance of female choice versus male–male competition in the evolution of sexually selected traits. *Behav. Ecol. Sociobiol.* **35**, 389–400.
- Berger, J. Induced abortion and social factors in wild horses. *Nature* **303**, 59–61 (1983).
- Clutton-Brock, T. & Langley, P. Persistent courtship reduces male and female longevity in captive tsetse flies *Glossina moritans* Westwood (Diptera: Glossinidae). *Behav. Ecol.* **8**, 392–395 (1996).
- Hoyt, D. F. Practical methods of estimating volume and fresh weight of bird eggs. *Auk* **96**, 73–77 (1979).
- Elder, W. H. & Weller, M. W. Duration of fertility in the domestic mallard hen after isolation from the drake. *J. Wild. Manag.* **18**, 495–502 (1954).

Acknowledgements

We thank the Duke and Duchess of Devonshire for kindly allowing us to conduct the study

at Chatsworth estate, the Chatsworth keepers for all their support and J. Shutt, M. Fowlie, D. Ross and M. Wilson for invaluable help with fieldwork. We would also like to thank B. Hatchwell, T. Birkhead, R. Johnstone, N. Davies, B. Appleby and A. Radford for help, advice and comments. The study was funded by NERC, UK.

Correspondence and requests should be addressed to E.C.
(e-mail: ejac3@hermes.cam.ac.uk).

Macaque monkeys categorize images by their ordinal number

Tanya Orlov, Volodya Yakovlev, Shaul Hochstein & Ehud Zohary

Department of Neurobiology, Silberman Institute of Life Sciences,
Hebrew University, Jerusalem, 91904, Israel

The recall of a list of items in a serial order is a basic cognitive skill¹. However, it is unknown whether a list of arbitrary items is remembered by associations between sequential items^{2,3} or by associations between each item and its ordinal position⁴. Here, to study the nonverbal strategies used for such memory tasks^{5–9}, we trained three macaque monkeys on a delayed sequence recall task. Thirty abstract images, divided into ten triplets, were presented repeatedly in fixed temporal order. On each trial the monkeys viewed three sequentially presented sample stimuli, followed by a test stimulus consisting of the same three images and a distractor image (chosen randomly from the remaining 27). The task was to touch the three images in their original order without touching the distractor. The most common error was touching the distractor when it had the same ordinal number (in its own triplet) as the correct image. Thus, the monkeys' natural tendency was to categorize images by their ordinal number. Additional, secondary strategies were used eventually to avoid the distractor images. These included memory of the sample images (working memory) and associations between sequence triplet members. Thus, monkeys use multiple mnemonic strategies according to their innate tendencies and the requirements of the task.

A basic question in the study of serial memory concerns the nature of the mental representation that allows retrieval of list items. Is list memory better characterized as a set of associations between adjacent items ('chaining'²) or even nonadjacent items³, or as a stored pattern of symbolic associations between each item and its ordinal position⁴? Humans can encode list items as words ('naming'), acquiring associations among names or between names and ordinal positions, and are very experienced with sequence tasks (numbers, alphabet, musical scale) and the abstract categories 'first', 'second' and so on. However, monkeys and even pigeons can reproduce ordered lists of arbitrary stimuli^{1,5–7,10–15}. What preverbal mnemonic strategies are used to represent lists?

In principle, multi-image working memory could be used to recall the identity of a stimulus and its temporal position in a sequence. This strategy is advantageous for generalization: it is as successful with new as with familiar stimuli, and is also successful with randomly ordered images. Monkeys can report whether a probe item was included in a previously presented list of up to 20 items almost as accurately as humans¹⁶. However, remembering the temporal position of an image is much more demanding. When monkeys were asked to choose which of two images appeared earlier in a list of five arbitrarily chosen items, their average accuracy was only ~ 73% (ref. 17).

When images are presented in fixed temporal order, subjects can use long-term memory strategies instead of working memory. One strategy is to generate an association between adjacent images in the

fixed sequence. Monkeys are highly skilled at generating such paired associations^{18–20}. Using a series of pair associations, a chain of images can be recalled. In fact, monkeys trained on a series of images successfully report the order of a random pair from the fixed sequence, even using non-sequential pairs^{5–7}. Similarly, monkeys trained on a sequence of pairs (A–B, B–C and so on) successfully report the order of any pair^{8,9}. In both cases analysis of response times indicates that they may use an internal serial recap of the list, perhaps on the basis of serial pair associations^{7,8,14}.

Another possible strategy is to memorize the ordinal position of each image. In one study, monkeys were trained on four nonverbal lists, each containing four novel photographs of natural objects, using the successive-phase method¹⁵. The task was to touch the simultaneously presented images in the correct order (A1–A2–A3–A4, B1–B2–B3–B4, C1–C2–C3–C4, D1–D2–D3–D4). When the monkeys had mastered this task, the items were shuffled, taking one item from each list, so that in two derived lists the ordinal number of the items was maintained (for example, A1–D2–C3–B4) whereas in two others it was not (for example, B3–A1–D4–C2). Lists with maintained ordinal position were acquired rapidly and virtually without error, whereas derived lists in which the ordinal position was changed were as difficult to learn as new lists. This pattern of transfer to derived lists implies that the monkeys originally acquired some knowledge about each item's ordinal position, rather than simply generating a chain of serial pair associations for each list of items.

Here we created an experimental paradigm in which monkeys could use all three of these strategies. This allowed us to assess which strategy is the most natural, being the first to be learned, and which is dominant, having the biggest effect on performance. The monkeys were trained on a task requiring them to report a list of three images in the order of their previous presentation. We used thirty fractal images, presented in fixed temporal order. These were divided into ten constant non-overlapping triplets (Fig. 1a). Each trial consisted of a triplet of three sample images, with each image

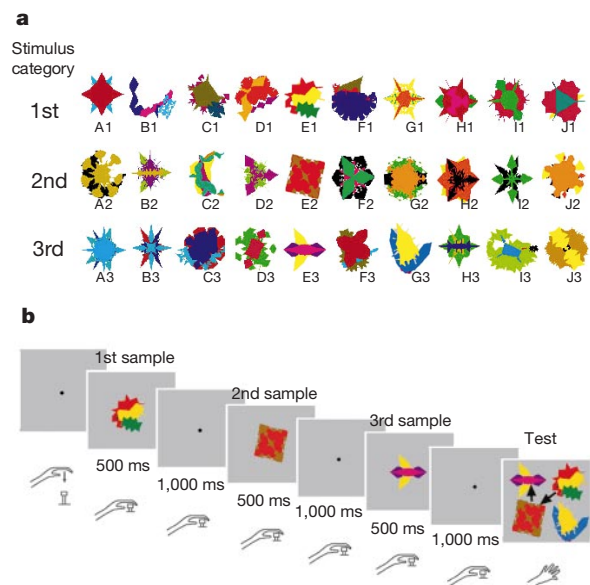


Figure 1 Image set and the basic delayed sequence recall task. **a**, A set of 30 fractal images were presented in a fixed temporal order using 10 constant non-overlapping triplets (columns). Each image has a fixed ordinal position in its triplet (1st, 2nd and 3rd) and the triplets were presented in a fixed order. The images were divided into three different categories (rows) according to their fixed ordinal positions in the triplets. **b**, Example of the sequence of events in a trial. Arrows at the 'Test' period mark the correct touching order.

having a fixed ordinal position in its triplet (first, second or third). During each trial, a single triplet was displayed twice: first, sequentially, during the sample sequence and then, simultaneously, during the test presentation. A distractor image, randomly chosen from the remaining 27 images, was added to the trial test presentation (Fig. 1b). The addition of this distractor image was essential to probe the monkeys' mnemonic strategies.

To study the nature of the routines that the monkeys used while learning the task, we analysed the distribution of their choices, taking into account the category of the distractor added to each test display. Note that within its own triplet the distractor had a fixed ordinal position.

Figure 2 shows the development of correct and erroneous (distractor) choices made during the first, second and third touch. The data are split between cases in which the distractor was from the same ordinal category as the correct choice image ('same' condition) and cases in which the distractor was from a different

ordinal category ('different' condition). In both conditions the monkeys learned to avoid touching the images belonging to the wrong ordinal category. Therefore, they rapidly learned to touch the correct image in the different condition, as it was the only one from its own category. (Monkey J had much less experience in the shaping stage, and therefore initially tended to make repeated choices of the same images in his second or third touches before extinguishing this behaviour.) Images belonging to the first and third ordinal categories were learned faster than the second ordinal category, suggesting primacy and recency effects^{16,21}. On the other hand, when the distractor was from the same ordinal category as the correct choice image, the monkeys erroneously touched the distractor almost as often as the correct image for a long period (graphs coloured yellow). They learned to discriminate between the distractor and the correct image much later in the same condition than in the different condition.

This dependence of performance on distractor category indicates that the monkeys' natural strategy was the generation of mnemonic representations of distinct image categories according to their ordinal positions within the triplets (first, second and third). Apparently, in the early stages of learning, the monkeys did not form associations between temporally adjacent images in each triplet, so they could not distinguish in this way between the distractor and the image that belonged to the current triplet. Furthermore, they did not perform the task by recalling which images were included in the sample sequence. Only later did the monkeys learn to use additional mnemonic strategies (associations between triplet members and working memory) to discriminate between the correct image and the distractor. To study quantitatively the role of alternative memory strategies in final task performance, we tested the monkeys on additional task paradigms when they reached asymptotic performance on the basic task (Table 1).

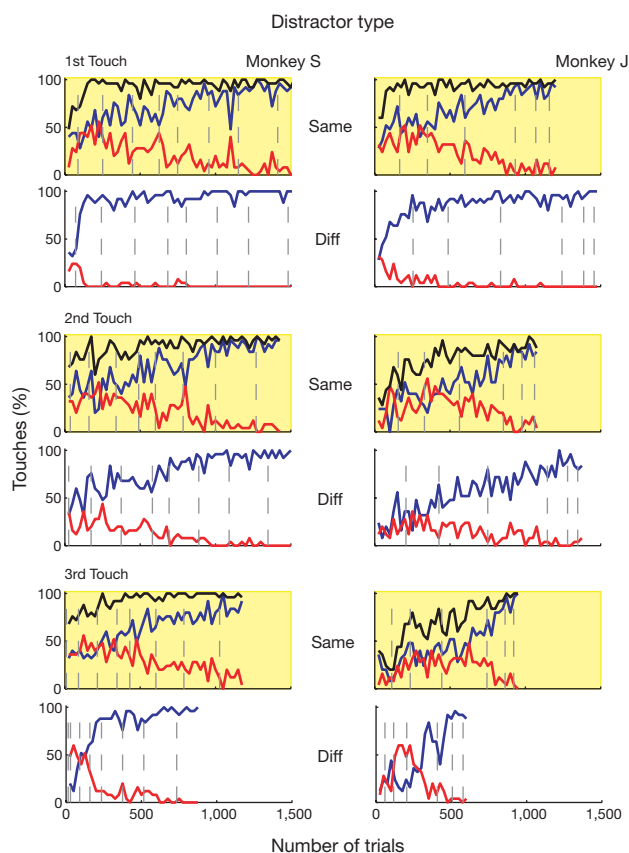


Figure 2 Basic task-learning dynamics. The proportion of choices (images touched) for the first, second and third touch of the trial, as a function of the number of trials, for monkeys S and J (each point represents data averaged over 25 trials). Correct touches are depicted in blue, choices of the distractor are in red. When the distractor was from the same ordinal category as the correct image (Same), the sum of these touches (that is, touches of the correct category) is shown in black. The other two choices (not shown here) complete the count to 100%. The distractor category type was either the same as that of the correct image (Same; yellow background), or from a different ordinal category than the correct image (Diff; distractors from third category for the first and second touch, and from first category for third touch are shown here). Performance improved rapidly for the 'different' condition, and only later for the 'same' condition. Monkey S chose the images from the correct category in its second and third touch earlier than monkey J, owing to greater experience in avoiding repeated touches in the shaping phase. Vertical dashed lines correspond to the number of completed trials every five sessions. The number of trials per session decreases monotonically as a function of the touch number because a wrong touch ended the trial.

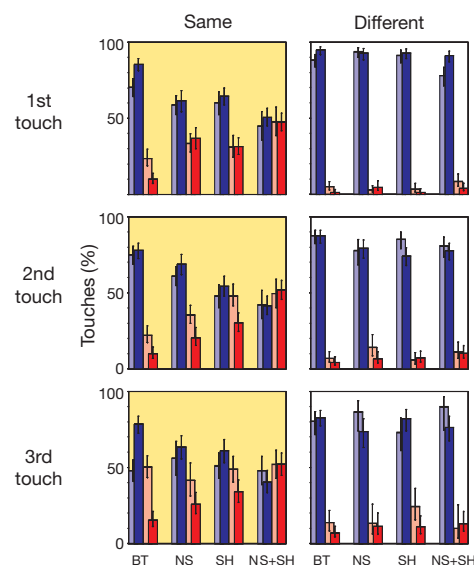
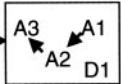
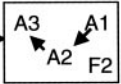
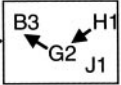


Figure 3 Performance in different versions of the task. Touch choices of the correct (blue bars) and distractor (red) images are shown for monkeys J (dark bars) and R (light bars). Task variants are BT: basic task, NS: no sample, SH: shuffled within categories and NS+SH: no samples and images shuffled within categories (Table 1). Cases in which the distractor was from the same category as the correct image and from a different category (as in Fig. 2) are shown in the left and right columns, respectively. The choices in the first, second and third touches are shown in the top, middle and bottom rows, respectively. Error bars denote a 95% confidence interval for the touch proportions. Note almost flawless performance in the 'different' condition when temporal order categorization suffices for correct response, compared to performance in the 'same' condition, which relies on the additional strategies available in each task version.

Table 1 Summary of the task versions and strategies available for each

Task version	Sample stimuli			Test stimulus	Strategies used	
	1st	2nd	3rd		Distractor from same category as correct stimulus	Distractor from different category than correct stimulus
					Ordinal number categorization and:	
BT (basic task)	A1	A2	A3		Working memory + associations	
NS (no samples)					Associations	Ordinal number categorization
SH (shuffled)	H1	G2	B3		Working memory	
NS+SH (no samples + shuffled)					None	

Besides the basic task (BT) we used three variants. NS: the initial sample stimuli were replaced by a blank, gray rectangle. SH: images were shuffled to form new triplets on each trial; shuffling maintained the ordinal category of each image, but destroyed the original fixed order of the set. The same shuffled images were presented in the test phase. NS+SH: combination of the two previous manipulations. Arrows in test stimulus denote correct response touch order. When the distractor category differed from that of the correct image, categorization by ordinal number sufficed for perfect performance. Otherwise, categorization narrowed the choice to the two images from the correct ordinal group, but recall of the samples (working memory) and/or associations among triplet members were required to discriminate between the correct image and the distractor.

In the first of these additional paradigms, we tested touch order without showing the initial sequence of stimuli on each trial (task NS) and found that the monkeys could still report the order of images²². When the distractor was from a different category from that to be touched, performance was nearly unaffected (Fig. 3, 'different' condition, NS; compare to basic task, BT). Only when the correct image and the distractor belonged to the same ordinal category was performance degraded—but not to chance level (50%) between images of the same category (Fig. 3, 'same' condition, NS; compare to BT). We observed a similar pattern of behaviour when we shuffled images within each category, disabling associations between sequential images (Fig. 3, task SH; compare 'same' and 'different' conditions). Finally, when we combined these two manipulations (NS+SH), the monkeys performed as well as in the basic task (BT) provided that the correct image belonged to a different category than the distractor (Fig. 3, 'different', NS+SH). Performance was reduced to chance level (50%) when the two images were from the same category.

Both monkeys achieved nearly perfect performance for the 'different' condition across task manipulations. This strongly indicates that the monkeys perform the task mainly by using retrieval of ordinal categories from long-term memory. Long-term associations between the images within each triplet (which are irrelevant in task SH but remain in task NS) and working memory (which is unavailable in task NS but potentially helpful in task SH) provide additional cues for differentiating between the distractor and the correct image when they belong to the same ordinal category.

Our results clearly indicate that monkeys' natural and dominant strategy for recalling ordered sequences of images is the retrieval of ordinal categories from long-term memory. An alternative explanation of our findings could be that the monkeys' choices reflect the relative salience of the different items. The items in the first ordinal category may have been more salient, because these were rewarded if touched first, whereas touching the others terminated the trial. Indeed, initially, the monkeys often touched the distractor from the first category for their second and third touches (see Fig. 2, third touch, 'different' condition). However, the monkeys did learn to

avoid touching the distractor in these cases. Thus, salience can not explain the results.

The monkeys also use working memory and recall of associations between triplet members, but only as secondary strategies recruited late in learning. This is most striking as either of these strategies, unlike ordinal number categorization, could have provided full information for solving the basic task. Categorization of the ordinal number of the images did not depend on creating chains between sequential images, as it was clearly established before the generation of associations between the triplet members (and concomitant discrimination between distractor and correct image). The ordinal position of each image was learned without acquiring the identity of its temporal neighbours. Monkeys naturally labelled the images as first, second and third, according to their temporal order, although they were probably unable to store the sequential order of a triplet of images in working memory.

The neural mechanisms underlying categorization of ordinal number are currently unknown. It is generally accepted that the medial temporal and frontal lobes are inherently involved in mnemonic function^{23,24}. Two possible cortical candidates for categorical long-term storage of visual image series are inferotemporal cortex and prefrontal cortex. Lesions in prefrontal cortex do not impair serial order behaviour that is based on retrieval from long-term memory^{25,26}, but the interaction between these two areas may be crucial for categorical memory²⁷. One study found that inferotemporal lesions degrade long-term memory of serial order²⁵. However, this could have been a result of impaired object recognition, not memory. Combining behavioural tests on the different variants of our task with either physiological recording or cortical lesions would allow a better assessment of the roles of these areas in long-term memory of image categories. □

Methods

Subjects and apparatus

Three adult male macaque monkeys, two *Macaca mulatta*, (J, 7.4 kg and R, 9.2 kg) and one *Macaca fascicularis* (S, 4.3 kg) were used in the experiments. Experiments were performed in accordance with NIH and Hebrew University guidelines for use of laboratory animals. The monkeys were housed individually and their water intake was regulated. We used

operant conditioning to train them on a memory task, giving a juice reward during training sessions. The monkeys sat in a primate chair 30 cm from a computer colour monitor equipped with a touch screen (31.5 cm × 26.5 cm). Trial events, stimulus presentation and data recording were computer controlled.

Behavioural shaping

The monkeys learned to press a lever following onset of a fixation spot and not to release it during presentation of three sample stimuli. At first, the stimuli in each sample presentation were chosen randomly (without repetitions) from the whole image set of 30 images. During the test phase of the trial, the same three stimuli were displayed on the screen simultaneously. This served as a 'go' signal for the monkeys to release the lever and touch the images on the screen. A touch of each image made in any order was rewarded, but repeated touches of the same image were considered as errors. During a brief final period of the shaping (monkey J: 100 trials; S: 400 trials; R: 700 trials) a fourth distractor image was added to the test stimulus, with the monkeys still rewarded irrespective of the order of touches, as long as they avoided repeated touches or touching the distractor.

In the tested behavioural task (described below), sample stimuli were shown as fixed ordered triplets and the image touches were rewarded only if they were made in the order of sample image presentation.

Stimuli

Thirty fractal images were presented repeatedly in a fixed temporal order, divided into ten constant non-overlapping triplets (A1, A2, A3; B1, B2, B3;...; J1, J2, J3; A1, A2, A3;...). Each trial consisted of a triplet of three sample stimuli, with each stimulus having a fixed ordinal position in its triplet (first, second or third). During each trial, a single triplet was displayed twice: first, sequentially, during the sample sequence and then, simultaneously, during the test presentation. A distractor stimulus, randomly chosen from the remaining 27 images, was added to the test presentation of the triplet (Fig. 1a, b).

Behavioural task

Monkeys were trained on a novel delayed sequence recall (DSR) task. The basic DSR task is shown schematically in Fig. 1b. Following presentation of a fixation spot on the screen, the monkey pressed a lever. This initiated presentation of a series of three sample stimuli (a triplet) in a fixed temporal order. Each sample image was presented at the centre of the screen for 500 ms with a 1 s inter-stimulus interval. Finally, the three sample images were presented together with a fourth, distractor stimulus at random positions on the screen. This test display served as a 'go' signal for the monkey to release the lever and touch the three sample images on the screen in the order of their previous presentation, without touching the distractor. Touching a wrong image ended the trial, whereas each correct touch was rewarded with juice. The third correct touch completing the trial was rewarded with a larger portion of juice. The number of trials per session varied considerably from session to session (between 17 and 269 trials). The monkeys were allowed to perform as many trials as they wished in a given session.

Task variations

After a period of prolonged training on the basic task, the monkeys were tested on additional task paradigms to study the role of different memory strategies on task performance:

- (1) Trials with no samples (task NS): In this task variant, we tested touch-order behaviour without presentation of the initial sample stimuli. During each of the 'sample' viewing times, the monkeys were presented with a grey rectangle, to maintain the temporal structure of the task.
- (2) Trials with stimuli shuffled between triplets but within categories (SH): In this variant the stimuli within each category were shuffled to form new triplets on each trial. Thus the shuffling maintained the ordinal category of each stimulus, but destroyed the original fixed order of the whole set. The same shuffled stimuli were presented in the test phase.
- (3) Trials with no samples and shuffled stimuli (NS+SH). We combined the two previous manipulations, shuffled the stimuli and displayed them only once, during the 'test' presentation. Reward was contingent only on touching the correct category.

Received 29 September 1999; accepted 4 January 2000.

1. Terrace, H. The phylogeny and ontogeny of serial memory: List learning by pigeons and monkeys. *Psychol. Sci.* **4**, 162–169 (1993).
2. Ebbinghaus, H. *Memory: A Contribution to Experimental Psychology* (Dover, New York, 1964 (1885)).
3. Young, R. K. The stimulus in serial verbal learning. *Am. J. Psychol.* **74**, 517–528 (1961).
4. Ebenholtz, S. in *The Psychology of Learning and Motivation* (ed. Bower, G.) 267–314 (Academic, New York, 1972).
5. D'Amato, M. & Colombo, M. Representation of serial order in monkeys (*Cebula apella*). *J. Exp. Psychol. Anim. Behav. Proc.* **14**, 131–139 (1988).
6. Terrace, H. Chunking by a pigeon in a serial learning task. *Nature* **325**, 149–151 (1987).
7. Swartz, K., Chen, S. & Terrace, H. Serial learning by rhesus monkeys: I. Acquisition and retention of multiple four-item lists. *J. Exp. Psychol. Anim. Behav. Proc.* **17**, 396–410 (1991).
8. McGonigle, B. & Chalmers, M. Monkeys are rational! *Q. J. Exp. Psychol.* **45**, 189–228 (1992).
9. Treichler, F. & van Tilburg, D. Concurrent conditional discrimination tests of transitive inference by macaque monkeys: List linking. *J. Exp. Psychol. Anim. Behav. Proc.* **22**, 105–117 (1996).
10. Straub, R. O. & Terrace, H. S. Generalization of serial learning in the pigeon. *Anim. Learn. Behav.* **9**, 454–468 (1981).
11. Terrace, H. S. A non-verbal organism's knowledge of ordinal position in a serial learning task. *J. Exp. Psychol. Anim. Behav. Proc.* **41**, 203–214 (1986).
12. Terrace, H., Chen, S. & Newman, A. Serial learning with a wild card by pigeons: Effect of list length.

J. Comp. Psychol. **109**, 162–172 (1995).

13. D'Amato, M. & Colombo, M. Serial learning with wild card items by monkeys (*Cebula apella*): implications for knowledge of ordinal position. *J. Comp. Psychol.* **103**, 252–263 (1989).
14. D'Amato, M. & Colombo, M. The symbolic distance effect in monkeys (*Cebula apella*). *Anim. Learn. Behav.* **18**, 133–140 (1990).
15. Chen, S., Swartz, K. B. & Terrace, H. S. Knowledge of the ordinal position of list items in rhesus monkeys. *Psychol. Sci.* **8**, 80–86 (1997).
16. Sands, S. F. & Wright, A. A. Serial probe recognition performance by a rhesus monkey and a human with 10- and 20-item lists. *J. Exp. Psychol. Anim. Behav. Proc.* **6**, 386–396 (1980).
17. Gower, E. Short-term memory for the temporal order of events in monkeys. *Behav. Brain Res.* **52**, 99–103 (1992).
18. Murray, E. I., Gaffan, D. & Mishkin, M. Neural substrates of visual stimulus-stimulus association in Rhesus monkeys. *J. Neurosci.* **13**, 4549–4561 (1993).
19. Sakai, K. & Miyashita, Y. Neural organization for the long-term memory of paired associates. *Nature* **354**, 152–155 (1991).
20. Naya, Y., Sakai, K. & Miyashita, Y. Activity of primate infero-temporal neurons related to a sought target in pair-association task. *Proc. Natl Acad. Sci. USA* **93**, 2664–2669 (1996).
21. Buffalo, B. & Gaffan, D. A primacy effect in monkeys when list position is relevant. *Q. J. Exp. Psychol.* **47B**, 353–369 (1994).
22. Farrar, D. N. Picture memory in the chimpanzee. *Percept. Motor Skills* **25**, 305–315 (1967).
23. Mishkin, M. A memory system in the monkey. *Phil. Trans. R. Soc. Lond.* **298**, 83–95 (1982).
24. Fuster, J. *The Prefrontal Cortex: Anatomy, Physiology and Neuropsychology of the Frontal Lobe* (Raven, New York, 1989).
25. Colombo, M., Eickhoff, A. E. & Gross, C. G. The effects of inferior temporal and dorsolateral frontal lesions on serial-order behavior and visual imagery in monkeys. *Cogn. Brain Res.* **1**, 211–217 (1993).
26. Petrides, M. Impairments on nonspatial self-ordered and externally ordered working memory tasks after lesions of the mid-dorsal part of the lateral frontal cortex in the monkey. *J. Neurosci.* **15**, 359–375 (1995).
27. Parker, A. & Gaffan, D. Memory after frontal/temporal disconnection in monkeys: conditional and non-conditional tasks, unilateral and bilateral frontal lesions. *Neuropsychologia* **36**, 259–271 (1998).

Acknowledgements

We thank D. Amit for insightful theoretical discussions and many practical suggestions during the course of these experiments, and S. Fusi, M. Dvorkin and O. Algaoy for computer assistance. This study was supported by grants from the Israel Science Foundation of the Academy of Arts and Sciences, the McDonnell-Pew Science Foundation and the US–Israel Binational Science Foundation (BSF). T.O. and V.Y. were supported in part by the Israel Ministry of Absorption.

Correspondence and requests for materials should be addressed to T.O.
(e-mail: tanyao@apollo.lsh.huji.ac.il).

Temporal patterns of human cortical activity reflect tone sequence structure

Aniruddh D. Patel* & Evan Balaban*

The Neurosciences Institute, 10,640 John Jay Hopkins Drive, San Diego, California 92121, USA

* These authors contributed equally to this work

Despite growing interest in temporal aspects of auditory neural processing^{1,2}, little is known about large-scale timing patterns of brain activity during the perception of auditory sequences³. This is partly because it has not been possible to distinguish stimulus-related activity from other, endogenous brain signals recorded by electrical or magnetic sensors. Here we use amplitude modulation of unfamiliar, ~1-minute-long tone sequences to label stimulus-related magnetoencephalographic neural activity in human subjects^{4–9}. We show that temporal patterns of activity recorded over particular brain regions track the pitch contour of tone sequences, with the accuracy of tracking increasing as tone sequences become more predictable in structure. In contrast, temporal synchronization between recording locations, particularly between sites over the left posterior hemisphere and the rest of the brain, is greatest when sequences have melody-like statistical properties^{10,11}, which may reflect the perceptual integration of local and global pitch patterns in melody-like

sequences¹². This method is particularly well suited to studying temporal neural correlates of complex auditory sequences (such as speech or music) which engage multiple brain areas as perception unfolds in time.

Five male human subjects heard sequences of pure tones while neuromagnetic signals were recorded using a 148-channel whole-head biomagnetometer (Biomagnetic Technologies). We took advantage of the fact that continuous amplitude modulation of tones results in detectable cortical activity at the modulation frequency, termed the 'steady-state response' (SSR)^{4–6}. The SSR is strongest when amplitude modulation is in the 40-Hz region⁷, and it may originate from two or more sources in each primary auditory cortex (Heschl's gyrus)^{8,9}. We applied a constant 41.5-Hz amplitude modulation (constant 'modulating frequency') to tone sequences (with varying 'carrier frequency') to identify cortical activity that was directly related to the stimulus. Listeners heard these stimuli as sequences of carrier-frequency pitches, and generated a robust SSR at 41.5 Hz whose timing (phase) varied as a function of the carrier frequency of the underlying tones (Fig. 1a–e).

Each subject heard 28 1-min tone sequences drawn from four structural categories. All categories used 415-ms-long pitches from

25 frequencies that were logarithmically equally spaced between 220 and 880 Hz (between musical A3 and A5 in semitone steps). We used a method of generating novel, unfamiliar stimuli in which different members of the same stimulus category were structurally unique, but shared common statistical properties¹³. The four structural categories differed in the relative balance of long-term pitch trends versus rapid point-to-point fluctuations (Fig. 1f–i, dashed black lines), which was controlled by specifying the slope of the pitch–time series power spectrum. In the first category, pitch–time patterns had a flat power spectrum corresponding to random variation from one pitch to the next. In the next two categories, pitch–time patterns had spectra with slopes of around $1/f$ and $1/f^2$, respectively, leading to two different 'fractal' patterns of constrained variation that more closely resembled musical melodies^{10,11,14,15}. The final category had deterministic pitch–time patterns consisting of linear step-wise motion up and down musical scales. Subjects were exposed to exemplars of the pitch patterns in a brief training session. Even though no information was given apart from examples and their category numbers (1 to 4), subjects had little difficulty forming perceptual categories and discriminating between the stimuli, as shown previously¹³. During the experiment, subjects identified the category number of each sequence after its presentation with few errors. With the exception of the scale pattern, every stimulus in the training session and experiment was unique, meaning that subjects discriminated the stimuli on the basis of statistical properties rather than by memorizing particular features.

For each trial in a given subject, data from each sensor were segmented into 2-s epochs and fast Fourier transformed to examine the magnitude and phase of the neuromagnetic signals in a 0.5-Hz band centred at 41.5 Hz. This gave energy (magnitude²)–time and phase–time series for each sensor. When the phase–time series were unwrapped and detrended (see Methods), recordings from particular sensor positions showed a marked resemblance to the contour of the pitch–time series heard by the subject (Fig. 1f–i, solid red lines). Energy–time series (the amplitude of the SSR response over time) never showed this pattern. For a given subject, sensor locations whose phase–time series showed a nominally significant positive correlation with the pitch–time series at the $P \leq 0.05$ level (one-tailed test) on >25% of all trials were designated 'tracking locations'. These locations had a consistent bilateral spatial distribution in all subjects, with a statistical tendency for their density to be greater on the right side across all subjects (Fig. 2a, 42 left compared with 83 right, χ^2 value relative to a symmetric distribution = 6.884, $P = 0.0102$, Fisher exact test¹⁶). When tracking performance was compared across the four stimulus categories at these locations, each subject showed highly significant differences in tracking between the categories (all $P < 0.0001$, Kruskal–Wallis analysis of variance (ANOVA)¹⁶). The correlation between neuromagnetic phase–time and stimulus pitch–time contours increased in the order random < $1/f < 1/f^2 < \text{scales}$ in all subjects (Fig. 2c). Across subjects, this pattern of increasing correlation with increasing stimulus structure was highly significant (Friedman two-way ANOVA¹⁶, $\chi^2 = 15.00$, d.f. = 3, $P = 0.0018$).

Energy changes in the neuromagnetic signal did not track stimulus structure. For each subject, baseline energy was computed in a 0.5-Hz band centred at 41.5 Hz for each channel over at least 50 2-s silent epochs recorded before the onset of stimulus presentations. Sensor locations where mean energy exceeded baseline by more than 1.96 s.d. on >25% of all trials were designated 'energy locations'. Energy and tracking locations overlapped in each subject, though one did not completely predict the other (Fig. 2b). No subject showed significant differences in energy between conditions at their energy locations (Kruskal–Wallis ANOVA, all $P > 0.25$, Fig. 2d). This suggests an aspect of the neural response that is actively related to the presence of an auditory stimulus but which does not vary with the structure of pitch–time sequences.

Our observations of cortical phase tracking of pitch sequences,

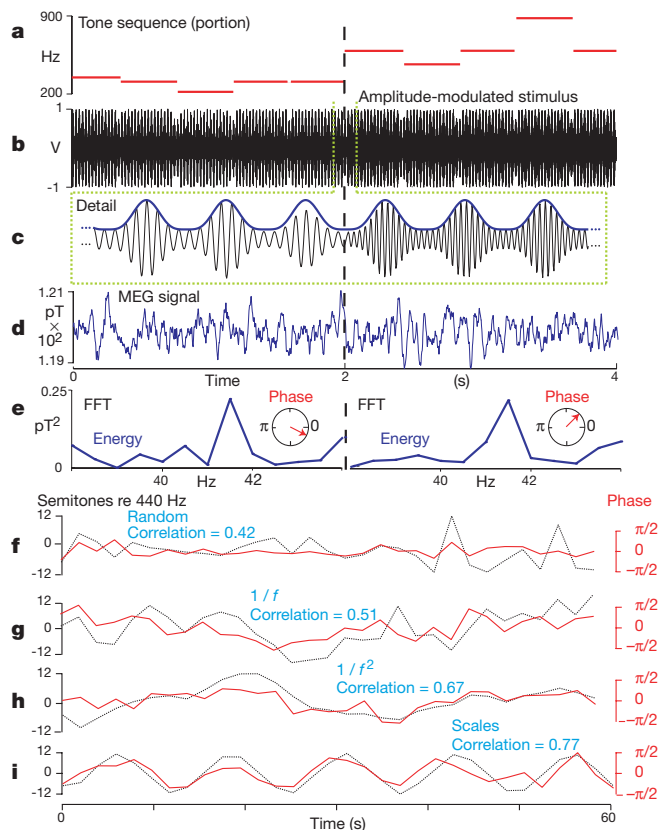


Figure 1 Phase tracking of tone sequences by the human brain. **a–d**, Stimulus and brain response over 4 s. **a**, Tone frequencies; **b**, stimulus waveform; **c**, waveform detail (150 ms), showing constant modulating frequency (purple) overlaid on changing carrier frequency (black); **d**, neural signal from one sensor (MEG, magnetoencephalography). **e**, Successive 2-s spectra of neural signal. The SSR has an energy peak at 41.5 Hz (purple), whose phase (red arrow) varies with carrier frequency. FFT, fast Fourier transform. **f–i**, Phase-tracking of individual tone sequences. Pitch–time contours (dotted black lines) illustrate the four different stimulus categories. Associated neuromagnetic phase–time series (solid red lines) from a single sensor during one trial in one subject were randomly drawn from the top 10% of sensor correlation values for each stimulus category. The correlation between the resampled pitch–time series and the neuromagnetic phase–time series is given in the inset to each graph. Tone sequences can be heard in the Supplementary information at <http://www.nature.com>.

consistent with earlier suggestions of a dependency between stimulus frequency and SSR phase⁴, have uncovered unexpectedly strong phase tracking over extended times, and a relationship between the quality of tracking and the statistical structure of pitch sequences. Although such tracking may depend upon frequency-dependent delays introduced in the cochlea⁴ and expanded on the way to the cortex¹⁷ these are unlikely to account for the modulation of tracking by pitch-sequence structure. This phenomenon may be due to a modulation of the number, phase synchrony or mean activity level of cells in the phase-tracking populations and/or a modulation of phase-noise in the 41.5-Hz frequency region by differential activity of cells which respond to the stimulus but are not involved in phase tracking. Modulation of these populations may originate from cortical and/or subcortical sites. Additional work is required to distinguish between these possibilities.

We also investigated the relationship between the statistical structure of pitch sequences and large-scale patterns of interaction among brain regions in the 41.5-Hz frequency band containing stimulus-related activity. Interactions were quantified in terms of phase coherence, a measure of temporal synchronization between pairs of sensors during a particular condition. To minimize the role of common neural sources, we included only sensor pairs >12 cm apart, as determined by regression analysis of coherence on inter-sensor distance for each subject¹⁸ (see Supplementary information). Phase coherences were computed for all such sensor pairs within each subject and condition after concatenating data for all seven trials. To remove the effects of absolute differences in sensor phase coherence and focus on changes between conditions, sensor interactions were classified by recording region (anterior left, anterior right, posterior left, posterior right) and type (tracking locations, non-tracking locations) and normalized within each subject (see Methods). The patterns of normalized phase coherence variation

among the four conditions were then combined across all subjects to investigate the significance of general and regional trends.

Across all subjects and brain areas a general trend emerged (Fig. 3a, Kruskal–Wallis ANOVA, d.f. = 3, $H = 499.6$, $P < 0.0001$, all post-hoc multiple comparisons significantly different at the 0.05 level). Random pitch sequences generated less coherence than the other, more patterned sequences, and $1/f^2$ sequences generated more coherence than deterministic scale patterns or $1/f$ sequences. Statistical research on Western musical pieces suggests that their pitch–time sequences have power spectra which are about $1/f^2$ (refs 10, 11), suggesting that music-like sequences generated the greatest synchronization between recording locations.

To investigate large-scale regional spatial patterns in coherence data, we segregated phase coherences according to the scheme shown in Fig. 3b. The general pattern in Fig. 3a reappears in the interactions between sensors above the posterior left brain hemisphere and those over the rest of the brain, which accounted for 45.6% of all sensor comparisons (Fig. 3c, Kruskal–Wallis ANOVA, $H = 173.4$, 33.7, 91.2, all $P < 0.0001$ for interactions vii, viii and ix, respectively; $1/f^2$ sequences show significantly greater coherence than the other conditions in post-hoc multiple comparisons at the 0.05 level). In contrast, in the remaining inter-regional comparisons (43.4% of comparisons) $1/f$ and $1/f^2$ sequences showed equivalent phase coherence (Fig. 3d, Kruskal–Wallis ANOVA, $H = 124.1$, 147.752, 157.5, all $P < 0.0001$ for interactions v, vi and x, respectively). Comparisons within the four regions (11% of all sensor comparisons) showed no consistent trends in their pattern of variation (Fig. 3e).

Imaging studies of music based on metabolic measures¹⁹ have largely implicated right fronto-temporal circuits in melodic processing. Studies of brain-damaged patients indicate that left superior temporal areas may be involved in the processing of local

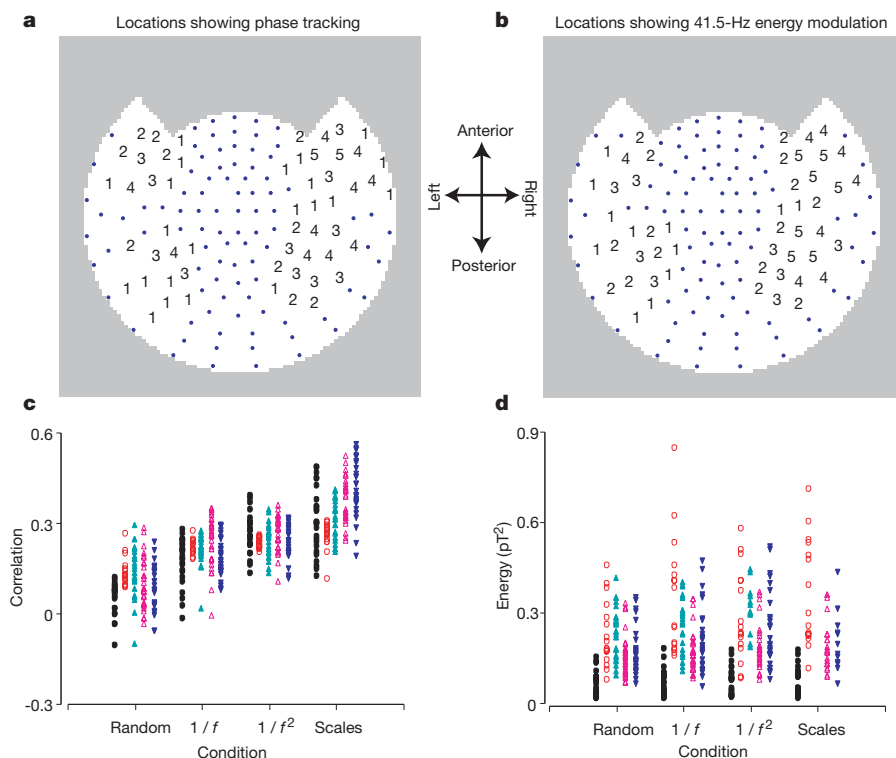


Figure 2 Topographic distribution of sensor locations. **a**, Locations where phase co-varied with tone frequency; **b**, locations where responses showed energy modulation at 41.5 Hz. Numbers indicate how many subjects (out of five) at each sensor location exceeded phase-tracking criteria (**a**) and energy-threshold criteria (**b**). Dots indicate locations that did not meet these criteria. **c**, Correlations between tone frequency and

response phase for the four types of sequence. Each data point represents the mean correlation over seven trials for one type of tone sequence for one subject at one phase-tracking sensor location. Each subject is shown with a single colour/shape combination. **d**, Mean 41.5-Hz energy at energy locations as a function of stimulus condition; symbols as in **c**.

melodic intervals in musical sequences, whereas right fronto-temporal circuits are involved in the perception of global pitch contour^{12,20}. Our data point to a potential neural correlate of the perceptual integration of local and global pitch information, and suggest that this integration is maximal for melody-like sequences. Dynamic imaging techniques can thus provide an important complement to metabolic imaging (such as functional magnetic resonance imaging or positron emission tomography) in the mapping of brain activity associated with music perception.

Auditory steady-state responses to amplitude-modulated sounds have been studied by a number of researchers interested in both the tonotopic organization of human auditory cortex^{5,21} and mechanisms of cortical signal generation^{6–8,22}. Such studies have typically used repetitive stimuli and extensive time-domain averaging to detect brain responses or to localize their sources. Our approach provides complementary information by elucidating dynamic temporal aspects of neural responses. This technique is especially suitable for the continuously changing sequences that are ubiquitous in speech and music, as well as for studying dynamic interactions between hearing and other sensory modalities such as vision. □

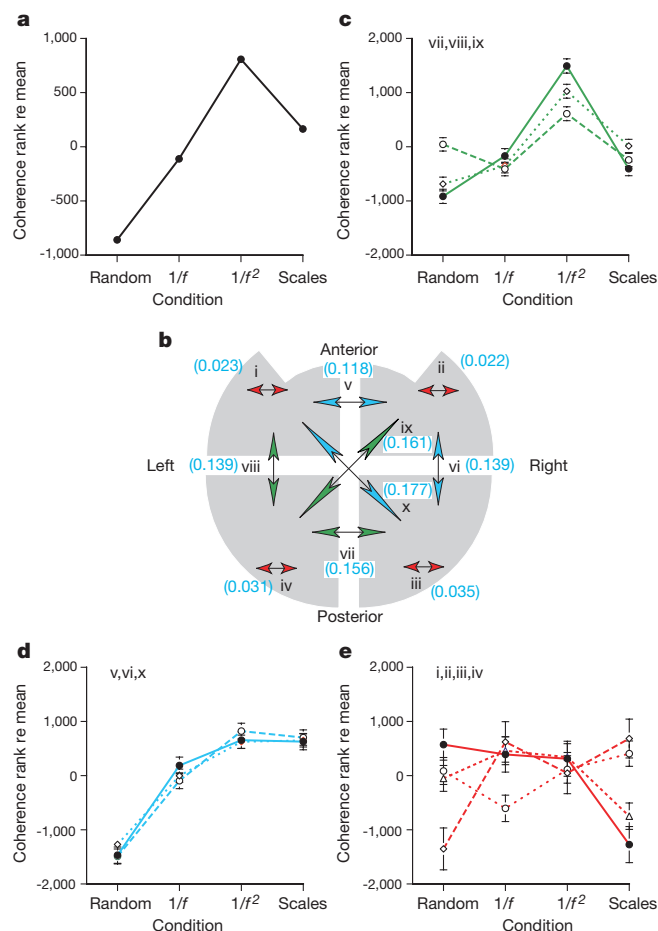


Figure 3 Changes in brain interactions among the four stimulus conditions. Graphs illustrate differences in mean ranks across conditions (error bars ± 1 s.e.m.; no bars, s.e.m. smaller than data point symbol). **a**, General pattern of coherence rank changes across all regions and subjects. **b**, Subdivision of brain regions into four quadrants (i, ii, iii, iv: sensor pairings within each quadrant; v–x: interregional interactions indicated by the neighbouring arrows). Numbers in parentheses: proportion of total sensor-pair interactions represented by each grouping. **c**, Inter-regional interactions involving sensors over the posterior left region: filled circle, vii; open circle, viii; open diamond, ix. **d**, Remaining inter-regional interactions: filled circle, v; open circle, vi; open diamond, x. **e**, Intra-regional interactions: filled circle, i; open diamond, ii; open triangle, iii; open circle, iv.

Methods

Stimulus preparation and presentation

Each tone sequence was created using SIGNAL (Engineering Design), by inverse-Fourier transforming complex spectra whose squared magnitudes declined with frequency at a rate of $1/f^b$, ($b = 0, 1.3$ and 2.1 for random, $1/f$ and $1/f^2$, respectively), and whose phases were random. For each sequence, 145 consecutive values were scaled and rounded to between ± 7 , corresponding to steps along Western musical scales (condition 4 consisted of linear ramps between ± 7 , with 150 total elements). Scale steps were converted to semitone steps (± 12 steps from 440 Hz) according to one of seven Western diatonic musical modes (ionian ('major scale'), dorian, phrygian, lydian, mixolydian, aeolian ('minor scale') and locrian; each mode appeared once in each condition). Semitone values were used to synthesize a tone sequence, with one 415-ms pure tone per frequency and no silences between tones (amplitude ± 1 V, last 20 ms of each tone 0.75 V). The entire tone sequence was amplitude-modulated at a rate of 41.5 Hz to a depth of 0.25 of its maximum amplitude using a $\cos^2$ envelope. Subjects were five right-handed males aged 30–36 who passed an audiometric test (GSI-65 Audiometer, Grayson-Stadler). Two of the subjects had studied music and the others had no special musical training.

Magnetoencephalographic recordings

Whole-head neuromagnetic signals were collected using a Magnes 2500WH MEG system (Biomagnetic Technologies) in a magnetically shielded room. This system provides 148 magnetometer coils (1 cm in diameter) spaced 3 cm apart on an approximately ellipsoidal surface located ~ 3 cm from the scalp surface. Stimuli were delivered over non-magnetic ER30 tube phones (Etymotic Research) at a comfortable level. The sampling rate for data acquisition was 508.63 Hz for one subject and 678.17 Hz for the others. Data were bandpass filtered from 1–100 Hz online during data acquisition.

Data preparation and analysis

Data preparation and analysis were carried out using MATLAB (MathWorks), SYSTAT (SPSS) and STATVIEW (SAS). Magnetoencephalographic data were fast-Fourier transformed in successive 2-s epochs (1,017-point transform for one subject and 1,356-point transform for the other four) and the magnitude and phase component of each Fourier coefficient were extracted at 41.5 Hz. Energy was estimated as the magnitude² in each epoch.

Phase–time series were unwrapped to correct for large jumps in radian phase and detrended (see Supplementary Information). To compare pitch–time and phase–time series, pitch–time series were down-sampled to match the Fourier transform rate of 1 value per 2 s. Correlations between down-sampled pitch–time series and phase–time series for each subject, trial and sensor were carried out after zero-meaning and root-mean-square amplitude normalization of both waveforms.

To analyse brain interactions in each subject, neuromagnetic data from the seven runs of each condition were concatenated end-to-end and Fourier transformed in 2-s epochs (giving ~ 217 epochs per sensor). Phase coherences were calculated as:

$$\gamma_{xy}^2 = \frac{|F_x/M_x \cdot cc(F_y/M_y)|^2}{|F_x/M_x \cdot cc(F_x/M_x)| |F_y/M_y \cdot cc(F_y/M_y)|}$$

where γ_{xy}^2 is the phase coherence between sensors x and y ; F_x and F_y are the complex Fourier-coefficient time series of sensors x and y at 41.5 Hz; M_x and M_y are the real-valued Fourier magnitude time series of sensors x and y ; $\cdot$ is the dot product operator; and cc is the complex conjugation operator²³. Dividing each Fourier coefficient by its magnitude ensures that the resulting measure is only sensitive to phase. A γ_{xy}^2 value of 1 indicates a constant phase difference between sensors x and y throughout a condition, whereas a γ_{xy}^2 value of 0 means a random phase relationship. The phase coherences of all sensors > 12 cm apart (see Supplementary information) in all four stimulus conditions were ranked from smallest to largest. Sensor locations were classed into four quadrants (anterior left, anterior right, posterior left and posterior right), giving 10 spatial comparison categories (four intra-regional comparisons and six inter-regional comparisons). Sensor comparisons were further subdivided into three types: tracking location channels compared with each other, tracking location channels compared with non-tracking location channels, and non-tracking location channels compared with each other, giving 30 categories of phase coherence comparison in the data set for each subject. As we were interested in exploring differences among the four conditions rather than absolute differences among the sensor categories, the coherence rankings within each category were corrected for level differences due to region and type by subtracting the mean value of the category within each subject. Trends were examined by combining these data across all subjects and examining the variation of the corrected coherence ranks with stimulus condition.

Non-parametric statistics were used in all analyses because of the highly non-normal distributions of correlations, coherences and energy values.

Received 22 September; accepted 6 December 1999.

- Cariani, P. A. & Delgutte, B. Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *J. Neurophysiol.* **76**, 1698–1716 (1996).
- deCharms, R. C. & Merzenich, M. Primary cortical representation of sounds by the coordination of action potential timing. *Nature* **381**, 610–613 (1996).
- Petsche, H., Richter, P., von Stein, A., Etlinger, S. & Filz, O. EEG coherence and musical thinking. *Musik Percept.* **11**, 117–152 (1993).
- Galambo, R., Makeig, S. & Talmachoff, P. J. A 40-Hz auditory potential recorded from the human scalp. *Proc. Natl Acad. Sci. USA* **78**, 2463–2467 (1981).
- Romani, G. L., Williamson, S. J. & Kaufman, L. Tonotopic organization of the human auditory cortex. *Science* **216**, 1339–1340 (1982).

6. Pantev, C. *et al.* Relationship of transient and steady-state auditory evoked fields. *Electroenceph. Clin. Neurophysiol.* **88**, 389–396 (1993).
7. Hari, R., Hämäläinen, M. & Joutsiniemi, S. -L. Neuromagnetic steady-state responses to auditory stimuli. *J. Acoust. Soc. Am.* **86**, 1033–1039 (1989).
8. Gutschalk, A. *et al.* Deconvolution of 40 Hz steady-state fields reveals two overlapping source activities of the human auditory cortex. *Clin. Neurophysiol.* **110**, 856–868 (1999).
9. Makeig, S., Jung, T.-P., Bell, A. J., Ghahremani, D. & Sejnowski, T. J. Blind separation of auditory event-related brain responses into independent components. *Proc. Natl Acad. Sci. USA* **94**, 10979–10984 (1997).
10. Nettheim, N. On the spectral analysis of melody. *Interface* **21**, 135–148 (1992).
11. Boon, J. P. & Decroly, O. Dynamical systems theory for music dynamics. *Chaos* **5**, 501–508 (1995).
12. Liégeois-Chauvel, C., Peretz, I., Babai, M., Laguitton, V. & Chauvel, P. Contribution of different cortical areas in the temporal lobes to music processing. *Brain* **121**, 1853–1867 (1998).
13. Schmuckler, M. A. & Gilden, D. L. Auditory perception of fractal contours. *J. Exp. Psychol.: Hum. Percept. Perform.* **19**, 641–660 (1993).
14. Voss, R. F. & Clarke, J. '1/f noise' in music and speech. *Nature* **258**, 317–318 (1975).
15. Voss, R. F. & Clarke, J. '1/f noise' in music: music from 1/f noise. *J. Acoust. Soc. Am.* **63**, 258–263 (1978).
16. Siegel, S. & Castellán, N. J. Jr *Nonparametric Statistics for the Behavioral Sciences* (McGraw Hill, New York, 1988).
17. Greenberg, S., Poeppel, D. & Roberts, T. in *Psychophysical and Physiological Advances in Hearing* (eds Palmer, A., Summerfield, Q., Rees, A. & Meddis, R.) 293–300 (Whurr, London, 1998).
18. Srinivasan, R., Russell, P., Edelman, G. & Tononi, G. Increased synchronization of neuromagnetic responses during conscious perception. *J. Neurosci.* **19**, 5435–5448 (1999).
19. Zatorre, R. J., Evans, A. C. & Meyer, E. Neural mechanisms underlying melodic perception and memory for pitch. *J. Neurosci.* **14**, 1908–1919 (1994).
20. Patel, A. D., Peretz, I., Tramo, M. & Labrecque, R. Processing prosodic and musical patterns: a neuropsychological investigation. *Brain Lang.* **61**, 123–144 (1998).
21. Pantev, C., Roberts, L. E., Elbert, T., Roß, B. & Wienbruch, C. Tonotopic organization of the sources of human auditory steady-state responses. *Hearing Res.* **101**, 62–74 (1996).
22. Ribary, U. *et al.* Magnetic field tomography of coherent thalamocortical 40-Hz oscillations in humans. *Proc. Natl Acad. Sci. USA* **88**, 11037–11041 (1991).
23. Bendat, J. S. & Piersol, A. G. *Engineering Applications of Correlation and Spectral Analysis* (Wiley, New York, 1993).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of *Nature*. For additional sound examples, see http://www.nsi.edu/users/patel/tone_sequences.

Acknowledgements

We thank L. Kurelowech for technical assistance, R. Srinivasan for advice and discussions, and S. Makeig, M. Kutas, T. Urbach and S. Hillyard for suggestions. This research was supported by the Neurosciences Research Foundation as part of its research program on Music and the Brain at The Neurosciences Institute.

Correspondence and requests for materials should be addressed to A.D.P. (e-mail: apatel@nsi.edu) or E.B. (e-mail: evan@nsi.edu).

Cannabinoids control spasticity and tremor in a multiple sclerosis model

David Baker^{*}, Gareth Pryce^{*}, J. Ludovic Croxford^{*}, Peter Brown[†], Roger G. Pertwee[‡], John W. Huffman[§] & Lorna Layward^{||}

^{*} Neuroinflammation Group, Department of Neurochemistry, Institute of Neurology, University College London, 1 Wakefield Street, London WC1N 1PJ and the Institute of Ophthalmology, UCL, London EC1V 9EL, UK

[†] The Medical Research Council Human Movement and Balance Unit, National Hospital for Neurology and Neurosurgery, Queen Square, London, WC1N 3BG, UK

[‡] Department of Biomedical Sciences, Institute of Medical Sciences, University of Aberdeen, Foresterhill, Aberdeen AB25 2ZD, UK

[§] Department of Chemistry, Clemson University, Clemson, South Carolina 29634-1905, USA

^{||} Multiple Sclerosis Society of Great Britain and Northern Ireland, 25 Effie Road, London SW6 1EE, UK

Chronic relapsing experimental allergic encephalomyelitis (CREAE) is an autoimmune model of multiple sclerosis¹. Although both these diseases are typified by relapsing-remitting paralytic episodes, after CREAE induction by sensitization to myelin antigens¹ Biozzi ABH mice also develop spasticity and tremor. These symptoms also occur during multiple sclerosis and are difficult to control. This has prompted some patients to find

alternative medicines, and to perceive benefit from cannabis use². Although this benefit has been backed up by small clinical studies, mainly with non-quantifiable outcomes^{3–7}, the value of cannabis use in multiple sclerosis remains anecdotal. Here we show that cannabinoid (CB) receptor agonism using *R*(+)-WIN 55,212, Δ^9 -tetrahydrocannabinol, methanandamide and JWH-133 (ref. 8) quantitatively ameliorated both tremor and spasticity in diseased mice. The exacerbation of these signs after antagonism of the CB₁ and CB₂ receptors, notably the CB₁ receptor, using SR141716A and SR144528 (ref. 8) indicate that the endogenous cannabinoid system may be tonically active in the control of tremor and spasticity. This provides a rationale for patients' indications of the therapeutic potential of cannabis in the control of the symptoms of multiple sclerosis², and provides a means of evaluating more selective cannabinoids in the future.

High doses of Δ^9 -tetrahydrocannabinol THC; (the major psychoactive component of cannabis) can inhibit the development of CREAE in rodents^{9,10}, but this has been attributed to immunosuppression preventing the conditions that lead to the development of paralysis, rather than to a direct effect on the paralysis itself^{9,10}. However, the action of cannabinoids on experimental spasticity and tremor remains uncertain because there have so far been no behavioural data on the effects of cannabinoids in animal models relevant to these symptoms of multiple sclerosis.

It is well established that repeated neurological insults occur during CREAE; these are associated with increasing primary demyelination and axonal loss in the central nervous system (CNS)¹. However, it was also evident that CREAE animals can develop additional clinical signs, including unilateral or bilateral fore- and hindlimb tremor (Fig. 1) and hindlimb spasticity (Fig. 2). These accumulate with disease duration and activity. Tremor was associated with voluntary limb movements, but in more severe cases it was persistent at a frequency of ~40 Hz (Fig. 1e). Although considerably faster than encountered in humans (~6 Hz), this frequency is consistent with tremor electromyography in mutant spastic (*Glr^bSpa*) mice¹¹. These animals develop episodes of rapid tremor and rigidity of the limb and trunk muscles¹². However, unlike the *Glr^bSpa* mouse, spasticity in CREAE mice need not be triggered by sudden disturbance¹². The effects of cannabis are mediated through the CB₁, CB₂ and putative CB₂-like receptors^{13,14}. CB₁ is predominant in the CNS and is the main target for psychoactivity, but it is also expressed at lower levels in many

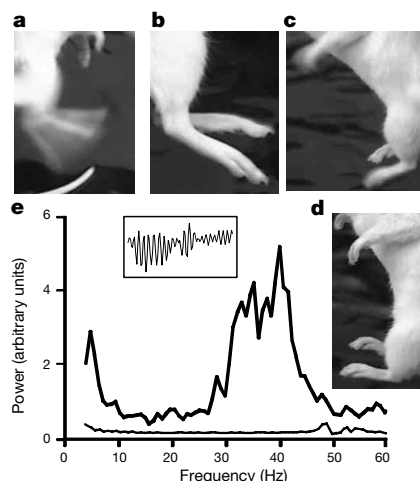


Figure 1 Cannabinoid receptor agonism inhibits tremor in autoimmune encephalomyelitis¹. Mice with hindlimb (a, b) or fore- and hindlimb (c, d) tremor both before (a, c) and after (b, d) treatment with 5 mg kg⁻¹ i.p. with *R*(+)-WIN 55,212. e, Power spectra of hindlimb tremors recorded with the foot suspended above a strain gauge before (thick line) and after (thin line) 5 mg kg⁻¹ i.p. *R*(+)-WIN 55,212 injection. Inset, snapshot of raw record over 0.5 s.

peripheral tissues. The CB₂ receptor is expressed at high levels on leucocytes, but there is also evidence for limited CB₂ receptor expression in mouse brain^{13,4}. The administration of a full CB₁ and CB₂ agonist, *R*(+)-WIN 55,212 (ref. 8), to post-relapse remission mice resulted in a rapid (within 1–10 min) amelioration of the frequency and amplitude of tremor in both the fore- and hindlimbs of CREA mice (Fig. 1). This was visually evident at 5 mg kg⁻¹ (Fig. 1a–d; *n* = 10/10) and 1 mg kg⁻¹ intraperitoneal (i.p.) (*n* = 6/6). In addition, Δ⁹-THC (10 mg kg⁻¹ intravenous (i.v.)) also ameliorated this response (*n* = 5/5). Tremor returned within hours after treatment. As Δ⁹-THC was observed to be relatively ineffective when injected intraperitoneally (i.p.), as seen in other studies¹⁰, all subsequent compounds were injected intravenously. Furthermore, as Δ⁹-THC is a partial CB₁ agonist but provides more limited CB₂ agonist activity, these results suggest that the effect on tremor is mainly mediated by the brain CB₁ receptor⁸.

Pretreatment (10 min) of animals with 5 mg kg⁻¹ i.v. of both selective CB₁ (SR141716A) (ref. 15) and CB₂ (SR144528) (ref. 16) receptor antagonists eliminated the capacity of 5 mg kg⁻¹ i.p. *R*(+)-WIN 55,212 to inhibit tremor (*n* = 5/5). However animals with residual paresis and mild spasticity became significantly more spastic after such CB receptor antagonism (Fig. 3). This was associated with uncontrolled leg crossing (Fig. 3c and d) and severe tail spasms. These showed gross curling which is atypical of post-remission animals, in which the tail generally hangs limply (Fig. 3e). Animals also show hindlimb extension (Fig. 3c), including a significant (*P* < 0.0001) increase in resistance to flexion (Fig. 3a, f). This was not observed in vehicle-treated controls (Fig. 3a). These signs were also not evident in similarly injected normal mice (*n* = 0/5) or normal-appearing pre-acute EAE animals (hindlimb resistance to flexion 0.159 ± 0.013N compared with 0.206 ± 0.022N in treated mice (*n* = 12 limbs, *P* > 0.05) and in animals with paresis/paralysis without evidence of spasticity (*n* = 0/5 treated with SR141716A and SR144528, *n* = 0/4 treated with SR141716A or SR144528 alone). When mildly spastic animals without tremor were injected with 5 mg kg⁻¹ i.v. CB₁ antagonist, not only did significant hindlimb (*P* < 0.001; Fig. 3a) and tail spasticity (*n* = 18/18, *P* < 0.001) develop compared with vehicle treated

controls (*n* = 0/6), but forelimb tremor also became evident in 3 out of 10 mice. This suggests a role for CB₁ in the control of tremor. After injection of 5 mg kg⁻¹ i.v. CB₂ antagonist, some animals (*n* = 10/14) seemed to show a mild increase in tail spasticity (*P* < 0.02) and showed a small but significant (*P* < 0.05) increase in resistance to hindlimb flexion (Fig. 3a). However, when the CB₂ antagonist was injected into animals previously made more spastic (*P* < 0.01) by CB₁ antagonism, spasticity increased significantly (*P* < 0.001) compared with animals treated with SR141716A alone, whereas this was resolved in animals treated with vehicle. This suggests that both CB receptors may control spasticity (Fig. 3f). However, it is possible that the effects of SR144528 could be mediated by CB₂-like (rather than CB₂) receptors as previously proposed¹⁷, or that at the dose used, SR144528 may have produced additional CB₁ antagonism because it has some limited capacity to bind to CB₁ (ref. 8). These observations may indicate the continual release of endogenous cannabinoid receptor agonists such as anandamide and 2-arachidonylglycerol which are present within the brain and exhibit neurotransmitter function¹⁸. Alternatively, or in addition, they may reflect the presence of precoupled, constitutively active cannabinoid receptors, as there is evidence that SR141716A and SR144528 are both inverse agonists that are capable of producing inverse cannabinimimetic effects by reducing the proportion of cannabinoid receptors that exist in a precoupled state^{8,15,16}. In comparison to some studies in which the antagonists affected the exogenous agonists¹⁷, the actions of the antagonists seen here were relatively short-lived (Fig. 3f). This may reflect the fact that the animals were attempting to compensate for the antagonist effect, and would be consistent with tonic control of the endogenous cannabinoid system. These data provide compelling evidence that CB receptors are involved in the control of spasticity in an environment of existing neurological damage, and that exogenous agonism may be beneficial.

Indeed, in mice with significant spasticity, 5 mg kg⁻¹ i.p. *R*(+)-WIN 55,212 reduced severity both visually (*n* = 7/7; Fig. 3g, h and i) and after assessment of resistance to hindlimb flexion (*P* < 0.001) (Fig. 3a and i). This was also evident with 2.5 mg kg⁻¹ i.p. *R*(+)-WIN 55,212 (Resistance of flexion of both limbs being reduced (*P* < 0.05) from 0.384 ± 0.096N to 0.276 ± 0.063N, *n* = 7, *P* < 0.05). Similar treatment with 5 mg kg⁻¹ i.p. of the inactive enantiomer *S*(-)-WIN 55,212 failed to significantly affect the spastic response (Fig. 3a). In contrast, 10 mg kg⁻¹ i.v. Δ⁹-THC and 5 mg kg⁻¹ i.v. methanandamide (CB₁-selective; *K*_i for CB₁ ≈ 20 nM and *K*_i for CB₂ ≈ 815 nM)⁸ induced a significant (*P* < 0.001) amelioration in spasticity (Fig. 3g). Coupled with the observations using SR141716A, this may suggest further that CB₁ is a main target for control of spasticity. Currently there are no compounds which are totally CB₁ or CB₂ receptor specific, but the lack of effect after 10 mg kg⁻¹ i.v. cannabidiol (main non-psychoactive component of cannabis. *K*_i for CB₁ = 4350 nM)⁸ suggested a subthreshold dose for CB₁ stimulation for treatment of spasticity. Using the CB₂-selective agonist JWH-133 (1.5 mg kg⁻¹ i.v. *K*_i for CB₁ ≈ 680 nM and *K*_i for CB₂ ≈ 3 nM)^{8,19} spasticity was reduced both 10 min (*P* < 0.05) and 30 min (*P* < 0.001) after injection at a time when 0.05 mg kg⁻¹ i.v. (dose selected to exhibit similar CB₁ activity to JWH-133) methanandamide was not active (Fig. 4). It is possible that sedative effects may have contributed (though CB₁ receptors) to cannabinoid-mediated effects in these assays, but there was no hypothermia, indicative of 'sedation' after JWH-133 administration (37.1 ± 0. °C (baseline), 37.2 ± 0.4 °C (10 min) 37.1 ± 0.2 °C (30 min)). That non-CB₁ receptors may also control spasticity is further indicated by the transient inhibition of spasticity with the endocannabinoid palmitoylethanolamide (Fig. 4). This compound has no significant affinity for CB₁ but may have activity for CB₂-like receptors⁸. The involvement of non-CB₁ receptors may be definitively resolved through the use of CB receptor subtype-specific compounds or CB-receptor-deficient mice.

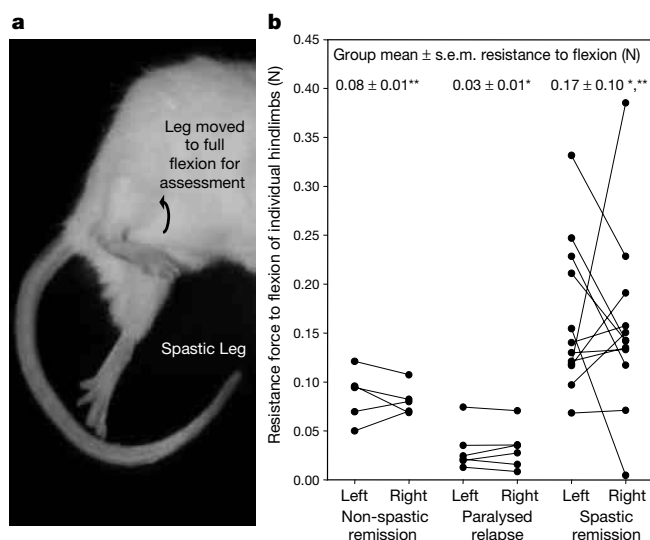


Figure 2 Spasticity develops in autoimmune encephalomyelitis¹. **a**, Spastic hindlimb showing full extension, including the digits. These were pressed against a strain gauge to measure the force required to bend the leg to full flexion. **b**, Increased resistance to flexion in post-relapse remission animals with spasticity (*n* = 12 mice) compared with age-matched mice without evidence of spasticity (*n* = 5 mice; asterisk, *P* < 0.001), or during active paralytic relapse episodes (*n* = 6; two asterisks, *P* < 0.001).

Spasticity in patients with multiple sclerosis can be very difficult to control despite the use of oral baclofen, dantrolene, diazepam and tizanidine, continuous intrathecal baclofen infusion, and selective injection of botulinum toxin²⁰. There is a need for more effective oral or systemic antispasticity agents. The hydrophobic nature of cannabinoids allows their rapid access to the CNS. Although the effects of chronic administration and dose dependency of CB receptor agonists on experimental spasticity remain to be

investigated further, the data presented here provide evidence for the rational assessment of cannabinoid derivatives in the control of spasticity and tremor in multiple sclerosis, in placebo-controlled trials. The observation that CB₁ appears to be the main therapeutic target suggests that it may be difficult to dissociate the full benefit from undesirable psychoactive elements using Δ^9 -THC or cannabis. It is also consistent with the unpleasant side effects experienced by some patients at the doses required for potential therapy by existing

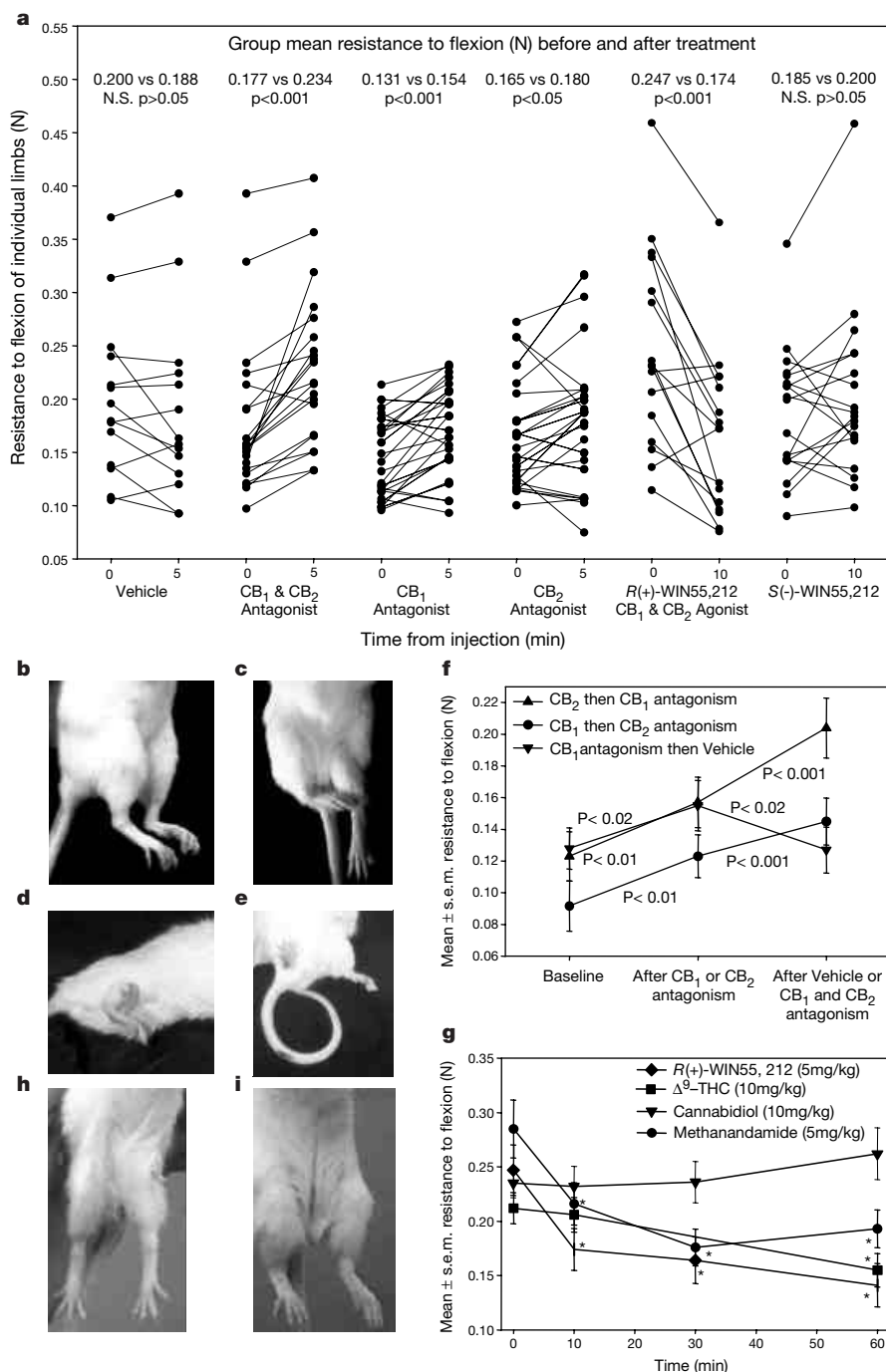


Figure 3 Control of spasticity by the cannabinoid system. **a**, Forces required to flex individual spastic hindlimbs against a strain gauge before and after injection with vehicle (0.1 ml i.v., $n = 14$), SR141716A (5 mg kg⁻¹ i.v., $n = 32$), SR144528 (5 mg kg⁻¹ i.v., $n = 32$), SR141716A and SR144528 ($n = 21$ limbs), R(+)-WIN 55,212 (5 mg kg⁻¹ i.p., $n = 16$) or S(-)-WIN 55,212 (5 mg kg⁻¹ i.p., $n = 19$). **b–e**, Cannabinoid receptor antagonism increased spasticity. Before (**b**) and after (**c**, **e**) SR141716A and SR144528 or after SR141716A (**d**) administration. **c**, **d**, Extension and crossing of limbs; **e**, spastic tail.

f, Resistance to flexion forces 5 min after SR141716A or SR144528 administration. 10 min later, mice were re-injected (5 mg kg⁻¹ i.v.) with either SR144528 ($n = 10$), vehicle ($n = 15$) or SR141716A ($n = 18$ limbs) and the resistance to flexion assessed after 5 min. **g**, Cannabinoid receptor agonism in spastic mice after either R(+)-WIN 55,212 ($n = 16$ limbs), Δ^9 -THC ($n = 18$), methanandamide ($n = 23$) or cannabidiol ($n = 22$). Asterisk, $P < 0.001$ compared with baseline. **h**, Spasticity was ameliorated (**i**) by treatment with R(+)-WIN 55,212.

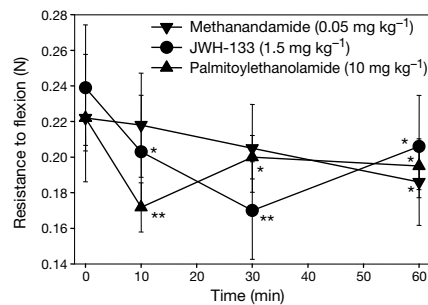


Figure 4 Treatment of spasticity in autoimmune encephalomyelitis¹ with non-CB₁ receptor agonists. Forces (mean $\pm$ s.e.m.) required to flex individual spastic hindlimbs against a strain gauge after i.v. injection with either low-dose methanandamide ($n = 9$ limbs), JWH-133 ($n = 9$) or palmitoylethanolamide ($n = 14$). Asterisk, $P < 0.05$; two asterisks, $P < 0.001$ compared with baseline.

cannabinoids³. The use of selective CB₂ agonists may provide some symptomatic benefit without significant psychoactive effects. Furthermore, it may be possible to upregulate endogenous produced cannabinoids¹⁸ to mediate therapeutic benefit. This CREAE model provides a means of evaluating and controlling the pathophysiology of spasticity in a chronic inflammatory environment relevant to the control of multiple sclerosis. □

Methods

Induction of CREAE

Biozzi ABH mice, bred at the Institute of Ophthalmology, were injected with 1 mg of mouse spinal cord homogenate emulsified in Freund's complete adjuvant on days 0 and 7 (ref. 1). Animals injected with CREAE, before the onset of acute phase CREAE¹ (usually occurring 15–20 days post inoculation (p.i.)) were used as normal CREAE controls. Paralyzed CREAE animals were selected during the acute phase or first relapse (typically occurring 34–45 days p.i.), and remission animals used for the assessment of tremor and spasticity were used after the second or third relapse 40–80 days p.i.).

Chemicals

R(+)-WIN 55,212, S(-)-WIN 55,212, Δ^9 -THC, methanandamide and cannabidiol were purchased from RBI/Sigma (Poole, UK). Palmitoylethanolamide was purchased from Tocris Cookson Ltd (Bristol, UK). SR141716A (ref. 15) and SR144528 (ref. 16) were supplied by M. Mossé and F. Barth (Sanofi Research, Montpellier, France). JWH-133 (3-(1'-1'-dimethylbutyl)-1-deoxy- Δ^8 -THC) was synthesised as described¹⁹. All compounds were dissolved at 0.5 mg ml⁻¹ in ethanol containing 1 mg ml⁻¹ Tween 80 (Sigma). The ethanol was removed by vacuum drying, and samples were reconstituted with phosphate buffered saline to a concentration of 2 mg ml⁻¹. Similar preparations without active drugs were used as vehicle controls. Suspensions (0.1 ml) were injected either i.v. or i.p. after CREAE induction.

Assessment of Clinical Signs

Spasticity and tremor were initially assessed by blinded analysis of video recordings. Digital images were sampled from video at 0.04 s. Signs of tail spasticity (flicking and curling) were assessed visually as being either present or absent. Spasticity was confirmed by assessing limb spasticity against a small purpose-built strain gauge. Limbs of animals without clinical evidence of spasticity (propensity to full extend the limb after tension on the leg) or the propensity to cross were not examined in drug studies. The analogue signal was amplified and digitally converted using an Amplicon card (Brighton, UK). This was captured using dacquire V10 software (D. Buckwell, MRC HMBU, Institute of Neurology) and analysed using Spike 2 software (Cambridge Electronic Design, UK). The hindlimbs were fully extended twice then moved to full flexion against the strain gauge. Each hindlimb was individually assessed by a blinded operator. The mean of 4–8 individual readings per limb was taken. Tremor frequency and severity were also recorded by holding the limb ~5 mm above the strain gauge. Tremor lead to the foot knocking the strain gauge. The strain gauge output was notch filtered at 50 Hz. The device had a resonance frequency of 95 Hz. The frequency of limb tremor was also confirmed using a lightweight unidirectional accelerometer (EGA XT-50, Entrain, UK) mounted over the foot.

Statistical Analysis

Results are expressed as means of individual feet or animals $\pm$ s.e.m. per group. The data were assessed using either a *t*-test, paired *t*-test for flexion data or nonparametric Mann-Whitney *U*-test using SigmaStat 2.0 software (Jandel Corp, San Rafael, California, USA).

Received 18 August 1999; accepted 20 January 2000.

- Baker, D. *et al.* Induction of chronic relapsing experimental allergic encephalomyelitis in Biozzi mice. *J. Neuroimmunol.* **28**, 261–270 (1990).
- Consroe, P., Musty, R., Rein, J., Tillery, W. & Pertwee, R. The perceived effects of smoked cannabis on patients with multiple sclerosis. *Eur. Neurol.* **38**, 44–48 (1997).
- Consroe, P. Cannabinoid systems as targets for the therapy of neurological disorders. *Neurobiol. Dis.* **5**, 534–551 (1998).
- Petro, D. J. & Ellenberger, C. Treatment of human spasticity with Δ^9 -tetrahydrocannabinol. *J. Clin. Pharmacol.* **21** (suppl.), 413–416 (1981).
- Clifford, D. B. Tetrahydrocannabinol for tremor in multiple sclerosis. *Ann. Neurol.* **13**, 669–671 (1983).
- Ungerleider, J. T., Andrysiak, T., Fairbanks, L., Ellison, G. W. & Myers, L. W. Δ^9 -THC in the treatment of spasticity associated with multiple sclerosis. *Adv. Alcohol Substance Abuse* **7**, 39–50 (1987).
- Martyn, C. N., Illis, L. S. & Thom, J. Nabilone in the treatment of multiple sclerosis. *Lancet* **345**, 579 (1995).
- Pertwee, R. G. Pharmacology of cannabinoid receptor ligands. *Curr. Med. Chem.* **6**, 635–664 (1999).
- Lyman, W. D., Sonett, J. R., Brosnan, C. F., Elkin, R. & Bornstein, M. B. Δ^9 -tetrahydrocannabinol: a novel treatment for experimental autoimmune encephalomyelitis. *J. Neuroimmunol.* **23**, 73–81 (1989).
- Wiggin, I. *et al.* Suppression of experimental autoimmune encephalomyelitis by cannabinoids. *Immunopharmacology* **28**, 209–214 (1994).
- Heller, A. H. & Hallet, M. Electrophysiological studies with the spastic mutant mouse. *Brain Res.* **234**, 299–308 (1982).
- Chai, C. K. Hereditary spasticity in mice. *J. Heredity* **52**, 241–243 (1961).
- Pertwee, R. G. Pharmacology of cannabinoid CB₁ and CB₂ receptors. *Pharmacol. Therapeut.* **74**, 129–180 (1997).
- Breivogel, C. S. & Childers, S. R. The functional neuroanatomy of brain cannabinoid receptors. *Neurobiol. Dis.* **5**, 417–431 (1998).
- Landsman, R. S., Burke, T. H., Consroe, P., Roeske, W. R. & Yamamura, H. I. SR141716A is an inverse agonist at the human cannabinoid CB₁ receptor. *Eur. J. Pharmacol.* **334**, R1–R2 (1997).
- Portier, M. *et al.* SR144528, an antagonist for the peripheral cannabinoid receptor that behaves as an inverse agonist. *J. Pharmacol. Exp. Ther.* **288**, 582–589 (1999).
- Calignano, A., La Rana, G., Giuffrida, A. & Piomelli, D. Control of pain initiation by endogenous cannabinoids. *Nature* **394**, 277–281 (1998).
- Giuffrida, A. *et al.* Dopamine activation of endogenous cannabinoid signalling in dorsal striatum. *Nature Neurosci.* **2**, 358–363 (1999).
- Huffman, J. W. *et al.* 3-(1'-1'-Dimethylbutyl)-1-deoxy- Δ^9 -THC and related compounds: synthesis of selective ligands for the CB₂ receptor. *Bioorg. Med. Chem.* **7**, 2905–2914 (1999).
- Noth, J. Trends in the pathophysiology and pharmacotherapy of spasticity. *J. Neurol.* **238**, 131–139 (1991).

Acknowledgements

The authors would like to thank the Multiple Sclerosis Society of Great Britain and Northern Ireland, the Medical Research Council, the National Institute on Drug Abuse and the Wellcome Trust for their financial support.

Correspondence and requests for materials should be addressed to D.B. (e-mail: D.Baker@ion.ucl.ac.uk).

Light acts directly on organs and cells in culture to set the vertebrate circadian clock

David Whitmore*, Nicholas S. Foulkes* & Paolo Sassone-Corsi

Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS-INSERM-ULP, 1 rue Laurent Fries, 67404 Illkirch Cédex, CU de Strasbourg, France

* These authors contributed equally to this work

The expression of clock genes in vertebrates is widespread and not restricted to classical clock structures^{1,2}. The expression of the *Clock* gene in zebrafish shows a strong circadian oscillation in many tissues *in vivo* and in culture, showing that endogenous oscillators exist in peripheral organs³. A defining feature of circadian clocks is that they can be set or entrained to local time, usually by the environmental light–dark cycle^{4,5}. An important question is whether peripheral oscillators are entrained to local time by signals from central pacemakers such as the eyes or are themselves directly light-responsive. Here we show that the

peripheral organ clocks of zebrafish are set by light–dark cycles in culture. We also show that a zebrafish-derived cell line contains a circadian oscillator, which is also directly light entrained.

The peripheral organs of zebrafish, in particular the heart and kidney, contain endogenous circadian oscillators, as measured by a rhythmic oscillation of *Clock* gene expression in culture³. To investigate how these clocks might be set to local time and the role played by light, we placed two groups of hearts and kidneys into culture, one group in constant darkness (DD) and another group on a light–dark cycle (LD) (Fig. 1). The cultured organs were exposed to a cycle of 14 hours of light and 10 hours of dark (14L:10D) over five days, matching exactly the cycle experienced by the fish before dissection. In DD, *Clock* levels continued to oscillate, although with a lower amplitude than that seen *in vivo*, peaking around zeitgeber time (ZT, where ZT 0 corresponds to 'lights on') 15 in the heart and ZT 21 in the kidney, as previously described³ (Fig. 1a, b, d and e). In the LD cycle, however, the amplitude and robustness of *Clock* oscillation in the heart were greater than in the DD group (Fig. 1c). Rhythm amplitude increased similarly in the kidney (Fig. 1f) where, in addition, the phase of peak expression appears to be shifted from ZT 21 under DD to ZT 15 under the LD cycle (Fig. 1d and e), the phase of the peak *in vivo*. Thus, light appears to influence *Clock* oscillation directly in these organs in culture.

To determine whether light can entrain the peripheral clocks in these organs we placed both hearts and kidneys on a reversed light–dark cycle (DL) and studied the timing of *Clock* oscillation (Figs 2 and 3). Two groups of organs were dissected from the same fish population on a 14L:10D cycle, and immediately placed either on the same 14L:10D cycle (LD) or on a reversed, 10D:14L cycle (DL).

In hearts maintained on the LD cycle, *Clock* expression continued to peak at ZT 15. However, in hearts on the reversed, DL cycle, the first *Clock* peak was delayed by 12 h on day 2 in culture, at the equivalent of ZT 3 on the original LD cycle. From day 2 onwards the peak and trough values of *Clock* expression were reversed between the two groups (Figs 2a, b and 3d), consistent with the hearts being entrained to the reversed, DL cycle. The kidney shows the same reversal of peak expression, with stable re-entrainment shown by the beginning of day 3 of the DL cycle (Fig. 3a, b and d). An alternative explanation for the reversal of the *Clock* rhythm could be that light directly suppresses *Clock* expression. This is unlikely, as light in the zebrafish has no significant acute effect on *Clock* expression (data not shown), similar to the situation for the mouse *Clock* gene^{6–8}. We also examined expression of another clock gene, *timeless*, under the same conditions as for *Clock*. Similarly to its mouse homologue^{9,10}, zebrafish *timeless* neither oscillates nor responds to a light–dark cycle (Figs 2c and 3c). To show that the circadian oscillator is stably entrained to the reversed light cycle we repeated the experiment, placing the cultures into DD for the final two days. The oscillation should continue with the newly established phase, whereas in the case of an acute effect of light, the peak would return to the original phase seen in the animal. In the heart, the 12-h shift is conserved in DD, showing that the circadian clock is entrained to the new LD cycle (Fig. 2d). Light can, therefore, directly influence the circadian oscillator in peripheral tissues in culture and set the phase of the clock.

We wondered whether changes in temperature induced by the periodic lighting could be responsible for entraining these oscillations^{11–15}. This seems unlikely, as the cultures were incubated in

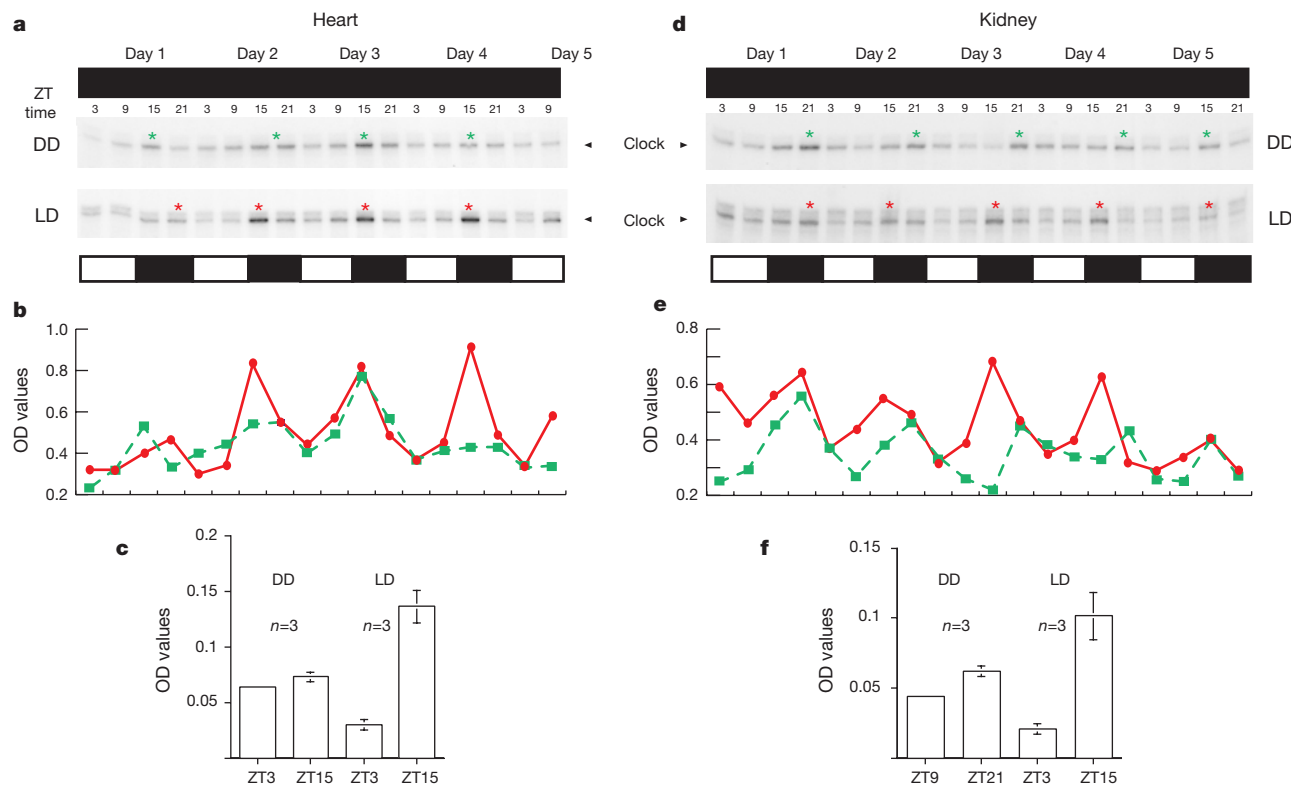


Figure 1 Exposure to light–dark cycles alters the pattern of *Clock* rhythmic expression in cultured zebrafish organs. **a**, Hearts were cultured for 5 days. At the indicated zeitgeber times (ZT) *Clock* messenger RNA was assayed by RPA (*Clock*-specific band indicated by arrowhead). Cultures were in constant darkness (DD) or a 14:10 light–dark cycle (LD). Asterisks, peaks of expression. **b**, Optical density (OD) measurements from **a**. The DD and LD values are green and red, respectively. **c**, Mean levels and s.e.m. at the trough (ZT 3) and peak (ZT 15) points, plotted for the fourth day in culture under DD and LD in additional

experiments. By unpaired Student's *t*-test analysis there was a significant oscillation under LD ($P < 0.005$) but not DD ($P > 0.1$). **d–f**, Equivalent analysis for kidney. The ZT 21 peak of expression under DD is shifted to ZT 15 in LD. Peak and trough values were measured at ZT 3 and ZT 15 under LD and ZT 9 and ZT 21 under DD. Differences were significant for both LD ($P < 0.01$) and DD ($P < 0.01$). Equivalent loading was verified as described³.

large volume water baths where there were no detectable temperature changes. We also placed cultured tissue on a precise 2 °C temperature cycle in DD, mimicking the light–dark cycles previously used. The circadian clock failed to entrain to these temperature cycles, but maintained the pattern of expression expected in DD. Therefore, a 2 °C temperature cycle, which far exceeds any recorded differences during our experiments, is insufficient to entrain these zebrafish clocks. Given these results, we believe that light acts directly on the circadian clock.

Evidence points to the existence of circadian oscillators even in immortalized cell lines^{16–18}. Circadian oscillations in gene expression have been reported in cell lines derived from the suprachiasmatic nucleus (SCN) and can be induced in fibroblast and hepatoma cell lines by serum shock followed by starvation^{16,17}. Therefore, we decided to examine a number of primary zebrafish cell lines^{19,20} to see whether they naturally contained a circadian oscillator that could also be influenced by exposure to light–dark cycles. One zebrafish embryonic cell line, PAC-2, showed constant, elevated levels of *Clock* expression in DD (Fig. 4a). However, when these cells were placed on to a 14L:10D cycle, an oscillation in *Clock* levels was apparent on the first day of the regime (Fig. 4b), with a relatively broad and low amplitude peak spread between ZT 9 and ZT 15. By the second day of the light cycle, the peak of expression had consolidated at ZT 15 and the amplitude of the rhythm had increased. When these cells were returned to DD the oscillation

continued for two cycles, but with a reduced amplitude and broadening of expression (Fig. 4c). By the third day in DD the oscillation was no longer apparent. This continuation of rhythmicity in DD following exposure to an LD cycle supports the hypothesis that these cells contain a clock that is entrained by the light–dark cycle, rather than a driving or masking effect of light on *Clock* expression.

The light entrainability of the PAC-2 cells indicates that they may contain a functional circadian oscillator that operates under steady-state culture conditions. However, the absence of significant rhythmic *Clock* expression under DD indicates that the LD cycle either synchronizes single-cell oscillators with a random phase distribution or initiates circadian oscillation that was not functional before light exposure. Single-cell imaging of cells expressing a fluorescent reporter gene under the control of a clock-regulated promoter may help to resolve this issue. Our results indicate that this cell line contains the photopigments and functional signal transduction cascades necessary for the light signal to reach the clock mechanism. The identity of these photopigments is not yet clear. Action spectra, as well as functional experiments, will be required to determine whether cryptochrome pigments are used, as appears to be the case in *Drosophila*, or whether novel opsins are the critical phototransducing molecules^{13,21}. In *Drosophila*, light-dependent entrainment of the circadian clock is thought to act through degradation of the Timeless protein as a consequence of sequestration by the cryptochrome photopigment^{22,23}. It will therefore also be

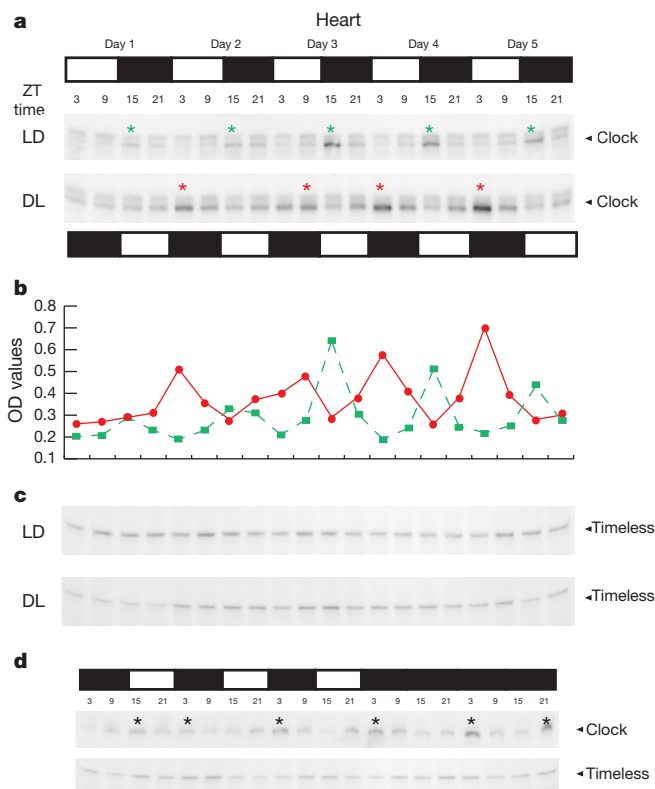


Figure 2 Clock-regulated rhythmic gene expression in the heart is entrained by reversed, DL cycles. **a**, RPA analysis of *Clock* transcript expression on LD or DL cycles for heart. Arrowheads, *Clock*-protected fragment; asterisks, times of peak expression. **b**, Optical density measurements from **a**. The rhythm of *Clock* expression is rapidly entrained to the reversed, DL cycle (red). For statistical analysis, see Fig. 3d. **c**, RPA analysis of zebrafish *timeless* expression in the same LD and DL samples. The non-oscillating, *timeless*-protected fragment is indicated. **d**, Hearts dissected from fish maintained in a normal, LD cycle were entrained to a reversed, DL cycle in L15 medium for 3 days in culture and then placed into DD for 2 days. In the upper gel, RPA of *Clock* expression shows that the re-entrained rhythm persists under DD. The lower gel shows non-oscillating *timeless* RPA analysis.

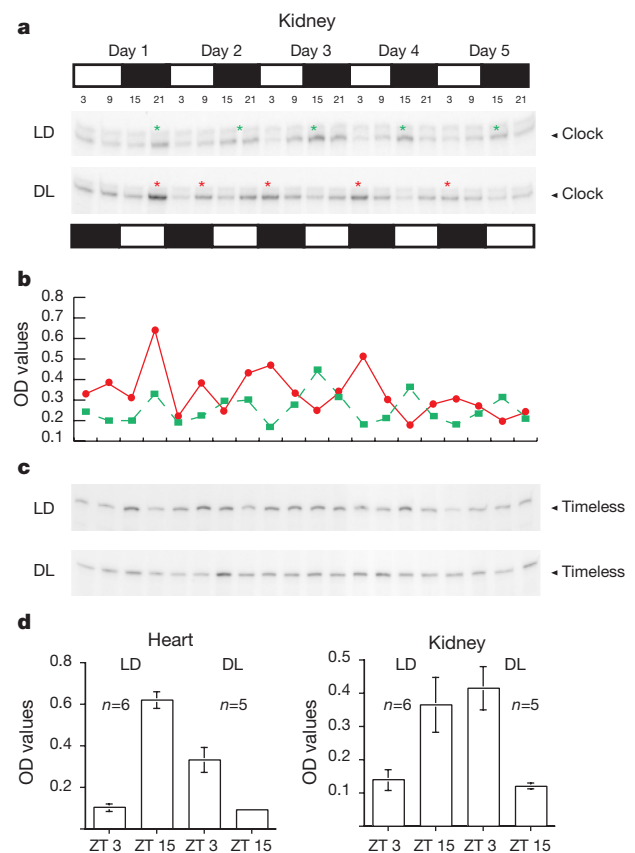


Figure 3 Entrainment of the *Clock* rhythm in the kidney by reversed, DL cycles. **a–c**, RPA analysis of *Clock* and *timeless* transcript expression in cultured kidneys, equivalent to that shown in Fig. 2a–c. **d**, Mean levels of *Clock* expression with s.e.m. at the trough (ZT 3) and peak (ZT 15) points plotted for hearts and kidneys on the third day in culture. Underlined ZT values for DL refer to the original light–dark cycle. By unpaired Student's *t*-test analysis differences between trough and peak expression were shown to be significant for both hearts and kidneys in both LD and DL conditions (hearts LD $P < 0.0001$, DL $P < 0.005$; kidney LD $P < 0.01$, DL $P < 0.005$).

of interest to study the effect of light on the expression of Timeless in our fish cell cultures.

The presence of a clock in an embryonic cell line may reflect the widespread presence of circadian clocks throughout the zebrafish body, where almost all tissues display rhythmic *Clock* expression (ref. 3 and unpublished data). Whether the clock is important during embryogenesis is unclear. However, rhythms in serotonin *N*-acetyltransferase (AANAT) and melatonin release have been described in embryos as early as 20–26 h after fertilization^{24,25}. Our description of direct light entrainability of the oscillators in hearts, kidneys and PAC-2 cells indicates that this might be a general property of clocks in a wide range of peripheral tissues. This is an unexpected result in a vertebrate where entrainment of the circadian clock has been thought to rely on the presence of circadian photoreceptors in the eye or pineal gland. The organization of the circadian system in the fish, therefore, seems to resemble that reported for *Drosophila*²⁶.

An important issue is the intensity of light that penetrates to particular organs. All of the experiments were performed initially in bright light of about 800 lux, but repeated with light intensities of approximately 200 lux with no apparent change in the speed of entrainment or amplitude of the resulting rhythm. In the context of the animal, it will be interesting to investigate whether other entraining signals such as melatonin or serum factors are involved^{16,27}. Entrainment to the shifted LD cycle appears to occur rapidly: hearts and kidneys reach the new phase by the beginning of the second and third days of DL, respectively. PAC-2 cells show rhythmicity on the first day of an LD cycle. The rapidity of this response may reflect the fact that in a population of cells with randomly distributed circadian phases, some of the cells require only small phase shifts to reach a correct phase relationship to the

LD cycle. The speed of clock entrainment in fish organs is comparable to that seen in rodent retina *in vitro* and in isolated *Drosophila* appendages, which also contain clocks and are directly light entrainable^{26,28}.

Our results add to the evidence that the circadian system in vertebrates exists as a decentralized collection of peripheral clocks. Individual organs have been previously shown to contain their own circadian oscillators^{3,29}, and now each tissue appears capable not only of detecting environmental light signals, but also using that information to set the phase of the clocks they contain. □

Methods

Fish

Zebrafish were raised from our own stocks and kept at 29 °C. They were fed twice daily and maintained under a 14-h day, 10-h night cycle. Adult fish (4 months old) were killed by rapid immersion in chilled water and decapitation. Dissections were performed under PBS using microdissection tools and a dissection microscope.

RNA analysis

RNA was extracted from zebrafish tissues and PAC-2 cell cultures using TRIzol reagent (Gibco BRL). We used a miniaturized RNase protection assay (RPA) using a zebrafish *Clock* complementary DNA-derived probe to assay *Clock* transcript expression as described³. A *timeless* probe for RPA was generated from zebrafish expressed sequence tag (EST) clone fa95g04. The 450-nucleotide probe corresponds to the carboxy-terminal region of mouse *timeless* between amino acids 921 and 1054, with which it shares 61% amino-acid similarity. Each point was prepared from a pool of six hearts or kidneys or one confluent 25-cm² flask of PAC-2 cells. RPA autoradiographs were scanned on an imaging densitometer (Biorad) and quantified using Molecular Analyst software (Biorad).

In vitro organ and cell culture

For *in vitro* organ cultures, freshly dissected tissue was placed in L15 medium supplemented with 15% fetal calf serum, 2 mM glutamine, gentamycin, streptomycin and penicillin. Organs were dissected from fish between ZT 9 and ZT 12 and placed directly into culture at 25 °C and atmospheric CO₂ concentration. A subline derived from the zebrafish embryonic cell line PAC-2 (ref. 19) was cultured as described above. Every 5 days, cells were trypsinized and replated at a 1:4 dilution. Cell stocks were maintained in non-illuminated incubators.

Lighting and temperature control

Culture flasks were completely immersed in a 60-litre, thermostatically controlled, circulating water bath. These cultures were illuminated using a tungsten light source connected to a programmable timer and a rheostat. The temperatures of cultures in the water bath were monitored using a thermocouple immersed in the culture medium, and light intensities were measured using a light meter (Gossen).

Received 25 August; accepted 14 December 1999.

- Whitmore, D., Sassone-Corsi, P. & Foulkes, N. S. PASTing together the mammalian clock. *Curr. Opin. Neurobiol.* **8**, 635–641 (1998).
- Reppert, S. M. A clockwork explosion! *Neuron* **21**, 1–4 (1998).
- Whitmore, D., Foulkes, N. S., Strähle, U. & Sassone-Corsi, P. Zebrafish Clock rhythmic expression reveals independent peripheral circadian pacemakers. *Nature Neurosci.* **1**, 701–707 (1998).
- Pittendrigh, C. S. Temporal organization: reflections of a Darwinian clock-watcher. *Annu. Rev. Physiol.* **55**, 17–54 (1993).
- Roenneberg, T. & Foster, R. G. Twilight times: light and the circadian system. *Photochem. Photobiol.* **66**, 549–561 (1997).
- Shearman, L. P., Zylka, M. J., Reppert, S. M. & Weaver, D. R. Expression of basic helix-loop-helix/PAS genes in the mouse suprachiasmatic nucleus. *Neuroscience* **89**, 387–397 (1999).
- Antoch, M. P. *et al.* Functional identification of the mouse circadian Clock gene by transgenic BAC rescue. *Cell* **89**, 655–667 (1997).
- King, D. P. *et al.* Positional cloning of the mouse circadian Clock gene. *Cell* **89**, 641–653 (1997).
- Zylka, M. J. *et al.* Molecular analysis of mammalian timeless. *Neuron* **21**, 1115–1122 (1998).
- Koike, N. *et al.* Identification of the mammalian homologues of the *Drosophila* timeless gene, Timeless1. *FEBS Lett.* **441**, 427–431 (1998).
- Barrett, R. K. & Takahashi, J. S. Temperature compensation and temperature entrainment of the chick pineal cell circadian clock. *J. Neurosci.* **15**, 5681–5692 (1995).
- Liu, Y., Merrow, M., Loros, J. J. & Dunlap, J. C. How temperature changes reset a circadian oscillator. *Science* **281**, 825–829 (1998).
- Stanewsky, R. *et al.* The cryb mutation identifies cryptochrome as a circadian photoreceptor in *Drosophila*. *Cell* **95**, 681–692 (1998).
- Zimmerman, W. F., Pittendrigh, C. S. & Pavlidis, T. Temperature compensation of the circadian oscillation in *Drosophila pseudoobscura* and its entrainment by temperature cycles. *J. Insect Physiol.* **14**, 669–684 (1968).
- Aschoff, J. & Tokura, H. Circadian activity rhythms in squirrel monkeys: entrainment by temperature cycles. *J. Biol. Rhythms* **1**, 91–99 (1986).
- Balsalobre, A., Damiola, F. & Schibler, U. A serum shock induces circadian gene expression in mammalian tissue culture cells. *Cell* **93**, 929–937 (1998).
- Earnest, D. J., Liang, F. Q., Ratcliff, M. & Cassone, V. M. Immortal time: circadian clock properties of rat suprachiasmatic cell lines. *Science* **283**, 693–695 (1999).

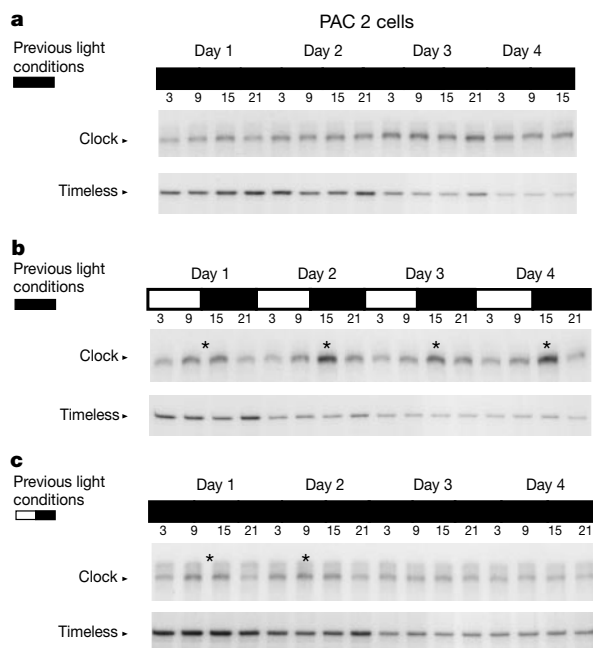


Figure 4 A zebrafish cell line contains a light-entrainable circadian clock. **a**, In cultures of the embryonic cell line PAC-2 maintained for 4 days in constant darkness no significant oscillation in *Clock* expression is detected. **b**, PAC-2 cultures previously maintained under DD were switched to an LD cycle (14 h light:10 h dark). A robust rhythm of *Clock* expression is first detected on day 1 and persists for the duration of the experiment. The peaks, around ZT 15, are indicated by asterisks. **c**, Cultures previously exposed for 5 days to a 14:10 LD cycle were returned to DD and *Clock* expression was monitored. The same pattern of rhythmic *Clock* expression persists for 2 days, but gradually declines during the last 2 days of the experiment. Below each *Clock* RPA analysis panel is shown the results of RPA for *timeless* expression in the same samples. The rhythm of expression visible for *Clock* is absent for *timeless*.

18. Earnest, D. J. *et al.* Establishment and characterization of adenoviral E1A immortalized cell lines derived from the rat suprachiasmatic nucleus. *J. Neurobiol.* **39**, 1–13 (1999).
19. Lin, S. *et al.* Integration and germ-line transmission of a pseudotyped retroviral vector in zebrafish. *Science* **265**, 666–669 (1994).
20. Peppelenbosch, M. P., Tertoolen, L. G., de Laat, S. W. & Zivkovic, D. Ionic responses to epidermal growth factor in zebrafish cells. *Exp. Cell Res.* **218**, 183–188 (1995).
21. Ishikawa, T. *et al.* DCRY is a *Drosophila* photoreceptor protein implicated in light entrainment of circadian rhythm. *Genes Cells* **1**, 57–63 (1999).
22. Zeng, H., Qian, Z., Meyers, M. P. & Rosbash, M. A light-entrainment mechanism for the *Drosophila* circadian clock. *Nature* **380**, 129–135 (1996).
23. Ceriani, M. F. *et al.* Light-dependent sequestration of TIMELESS by CRYPTOCHROME. *Science* **285**, 553–556 (1999).
24. Gothilf, Y. *et al.* Zebrafish serotonin N-acetyltransferase-2: Marker for development of pineal photoreceptors and circadian clock function. *Endocrinology* **140**, 4895–4903 (1999).
25. Kazimi, N. & Cahill, G. M. Development of a circadian melatonin rhythm in embryonic zebrafish. *Brain Res. Dev. Brain Res.* **117**, 47–52 (1999).
26. Plautz, J. D., Kaneko, M., Hall, J. C. & Kay, S. A. Independent photoreceptive circadian clocks throughout *Drosophila*. *Science* **278**, 1632–1635 (1997).
27. McArthur, A. J., Gillette, M. U. & Prosser, R. A. Melatonin directly resets the rat suprachiasmatic circadian clock *in vitro*. *Brain Res.* **565**, 158–161 (1991).
28. Tosini, G. & Menaker, M. Circadian rhythms in cultured mammalian retina. *Science* **272**, 419–421 (1996).
29. Cermakian, N., Whitmore, D., Foulkes, N. & Sassone-Corsi, P. *Proc. Natl Acad. Sci. USA* **97** (in the press).

Acknowledgements

We thank R. G. Foster, J. S. Takahashi, M. Menaker, S. Reppert, N. Cermakian, D. De Cesare, U. Strähle and P. Blader for discussions, advice and gifts of materials, and E. Heitz, D. Biellman, O. Nkundwa and N. Fisher for technical assistance. D.W. was supported by an EEC TMR fellowship. Our studies are funded by grants from CNRS, INSERM, CHUR, Rhône-Poulenc Rorer (Bioavenir), Fondation pour la Recherche Médicale and Association pour la Recherche sur le Cancer (P. S.-C.).

Correspondence and requests for materials should be addressed to P. S.-C. (e-mail: paolosc@titus.u-strasbg.fr).

Delayed activation of the paternal genome during seed development

Jean-Philippe Vienne-Calzada*[†], Ramamurthy Baskar*[†] & Ueli Grossniklaus*[†]

* Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA

Little is known about the timing of the maternal-to-zygotic transition during seed development in flowering plants. Because plant embryos can develop from somatic cells or microspores¹, maternal contributions are not considered to be crucial in early embryogenesis². Early-acting embryo-lethal mutants in *Arabidopsis*, including *emb30/gnom* which affects the first zygotic division^{3,4}, have fuelled the perception that both maternal and paternal genomes are active immediately after fertilization. Here we show that none of the paternally inherited alleles of 20 loci that we tested is expressed during early seed development in *Arabidopsis*. For genes that are expressed at later stages, the paternally inherited allele becomes active three to four days after fertilization. The genes that we tested are involved in various processes and distributed throughout the genome, indicating that most, if not all, of the paternal genome may be initially silenced. Our findings are corroborated by genetic studies showing that *emb30/gnom* has a maternal-effect phenotype that is paternally rescuable in addition to its zygotic lethality. Thus, contrary to previous interpretations, early embryo and endosperm development are mainly under maternal control.

In flowering plants, double fertilization involves two sperm cells: one fuses with the egg cell to form a diploid zygote; the second fuses

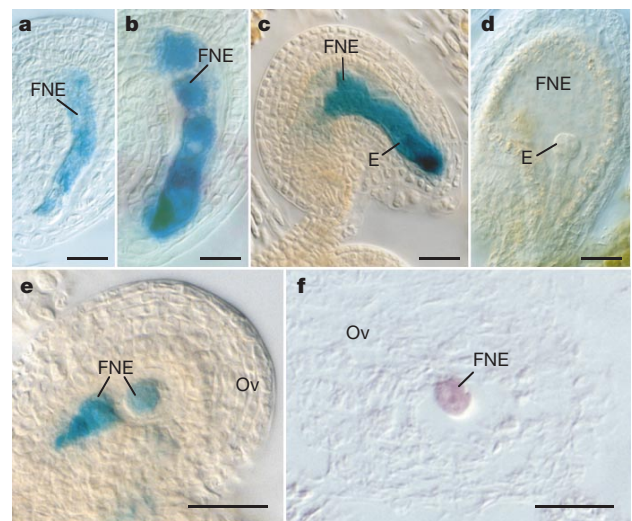


Figure 1 Silencing of paternally inherited genes during seed development in *Arabidopsis*. **a**, If an ET2612 female is crossed to a wild-type male, *GUS* expression is detected in the free nuclear endosperm 12 h.a.p. **b, c**, *F₁* seeds of the same cross show more intense *GUS* expression in embryo and endosperm 48 h.a.p. **d**, If ET2612 is crossed as the male to wild-type plants, *GUS* expression is not detectable in embryo and free nuclear endosperm 48 h.a.p. **e**, Transverse optical section through a seed of a self-pollinated ET2612 plant 48 h.a.p. **f**, *In situ* hybridization to mRNA of the putative basal transcription factor tagged in ET2612; the pattern of mRNA and *GUS* expression are identical 48 h.a.p. E, embryo; FNE, free nuclear endosperm; Ov, ovule. Scale bars, 17 μ m (**a, b**); 23 μ m (**c**); 45 μ m (**d**); 40 μ m (**e**).

with the binucleated central cell to give rise to the triploid primary endosperm nucleus⁵. Double fertilization triggers rapid proliferation of the endosperm and slow cell divisions of the zygote, which usually undergoes an asymmetrical division⁶. In *Arabidopsis*, as in most plant species, the primary endosperm nucleus undergoes divisions without cytokinesis, giving rise to a syncytium that eventually cellularizes⁷. In contrast to animals^{8–10}, the timing of transcriptional activation of the genome in the plant embryo and endosperm has not been intensively studied¹¹. The identification of a large group of early-acting embryo-lethal *Arabidopsis* mutants that segregate as sporophytic recessive traits¹² suggested that the activation of the zygotic genome occurs before the first division of the zygote; however, a spatial and temporal pattern for zygotic genome activation has not yet been determined.

We generated a library of enhancer detector and gene trap lines (transposants) that harbour *Ds* elements with a *uidA* reporter gene encoding β -glucuronidase (*GUS*) by using the system of Sundaresan *et al.*¹³. Screening for genes that act during ovule and early seed development in *Arabidopsis* (U.G. *et al.* unpublished data), we identified 19 transposants that show *GUS* expression in the developing embryo and/or endosperm after fertilization. *GUS* is expressed in the egg and/or central cell and persists for several rounds of cell division in either one or both fertilization products. To determine whether *GUS* expression was the result of transcription from one or both parental alleles in each of these lines, we performed reciprocal crosses between wild-type plants and the 19 transposants. When wild-type plants were used as male parents, the resulting *F₁* seeds showed *GUS* expression in a pattern identical to the one found in developing seeds resulting from self-pollination in all 19 lines (Fig. 1a–c). In contrast, if the transposants were used as male parents, *GUS* expression was absent from all *F₁* seeds and remained undetectable up to 80 hours after pollination (h.a.p.) (Fig. 1d). This indicates that the paternally inherited allele may not be expressed during early stages of embryo and endosperm development. To verify whether the pattern of *GUS* expression truly reflects the expression of genes neighbouring the insertion, we performed

[†] Present addresses: CINVESTAV-Irapuato, Departamento de Ingeniería Genética, Km 9.6 Carretera Irapuato-León, CP 36 500, Irapuato GTO, Mexico (J.-P.V.-C.); Friedrich Miescher Institute, Maulbeerstrasse 66, CH-4058 Basel, Switzerland (R.B., U.G.).

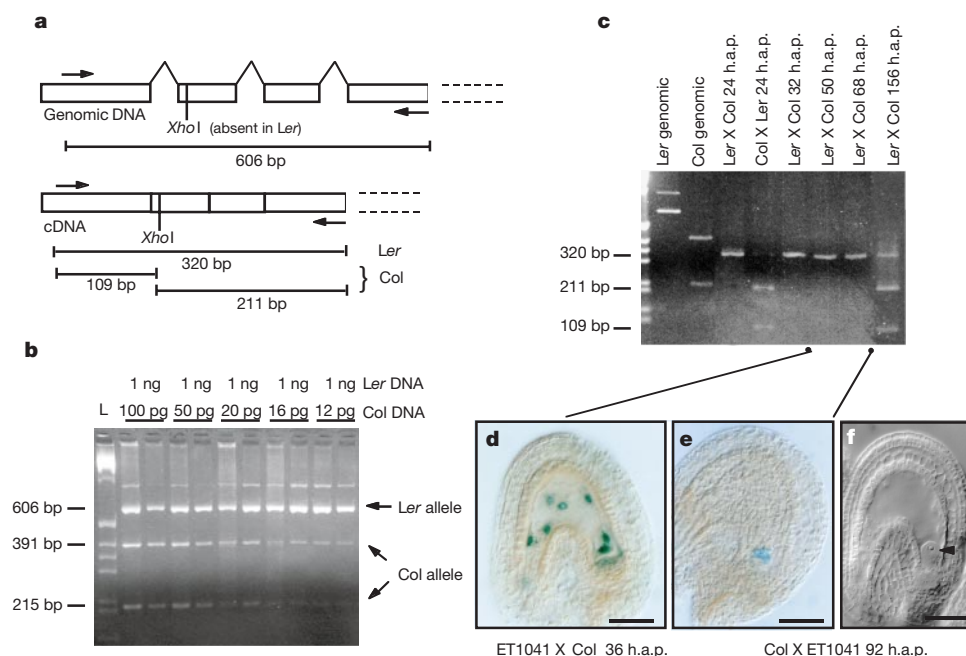


Figure 2 Allele-specific expression profile of *PRL* during early seed development. **a**, Schematic of *PRL* showing the position of exons (squared boxes), introns (lines) and the SNP in an *XhoI* site which is present in *Col* but absent in *Ler*. **b**, Allele-specific PCR analysis using titrated mixtures of genomic DNA each of which is represented twice. Both *PRL* alleles are consistently amplified throughout the dilution series. **c**, Allele-specific RT-PCR analysis of *PRL* in reciprocal crosses between *Col* and *Ler*. Only transcripts derived

in-situ hybridization with digoxigenin-labelled probes of a gene encoding a putative basal transcription factor that was tagged by ET2612 (see Table 1). Messenger RNA was detected in the developing endosperm (Fig. 1f) and the early globular embryo (data not shown) in a pattern identical to the one observed in *GUS* assays (Fig. 1e). Uniparental expression could be the consequence of transgene silencing specifically affecting the paternally inherited allele. Alternatively, it could indicate that our enhancer detection screen identified a subclass of genes that are transcribed maternally before fertilization and are subject to paternal silencing. A third possibility is that the paternally inherited genome remains silent during the first days after double fertilization, and that the activation of the paternal genome occurs after several rounds of cell division in both the embryo and endosperm.

To determine whether the absence of paternally derived *GUS* expression was due to sex-specific inactivation of the enhancer detector transgene or whether it reflected sex-specific silencing of endogenous genes, we designed a reverse transcription polymerase chain reaction (RT-PCR) assay that depends on single-nucleotide polymorphisms (SNPs) between Columbia (*Col*) and Landsberg *erecta* (*Ler*) ecotypes. Single-nucleotide polymorphisms in endonuclease restriction sites allow the distinction of transcripts derived from *Col* and *Ler* alleles. *PROLIFERA* (*PRL*) is an MCM2-3-5-like replication licensing factor that is required for the initiation of DNA replication¹⁴. We have identified a new enhancer detector insertion (ET1041) in the coding region of *PRL* (see Table 1), and reporter gene expression confirmed that *PRL* is expressed in the developing embryo and free nuclear endosperm. An SNP in the second exon of *PRL* creates a *XhoI* site in *Col* but not in *Ler* (Fig. 2a). To test the sensitivity of our assay, we amplified a region spanning exon 2 in titrated mixtures of genomic DNA. After digesting the PCR products with *XhoI*, we could consistently amplify both *PRL* alleles even when the amount of *Col* DNA used as a template was 80 times less than that of *Ler* DNA (Fig. 2b). We examined *PRL* expression by RT-PCR on RNA isolated 24, 32, 50, 68 and 156 h.a.p. from developing

from the maternal *PRL* allele are detected up to 68 h.a.p. **d**, If *PRL* enhancer detector ET1041 is used as the female in crosses to *Col*, *GUS* is detected in the endosperm 36 h.a.p. **e**, If *PRL* enhancer detector ET1041 is used as the male parent, *GUS* is detected in a small portion of the chalazal endosperm 92 h.a.p. **f**, Cleared seed showing the nodular cyst (arrowhead) in the chalazal endosperm. Scale bars, 53 μ m (**d**); 50 μ m (**e**); 55 μ m (**f**).

siliques derived from reciprocal crosses between *Ler* and *Col* plants. During the first three days after pollination (68 h.a.p.) only transcripts derived from the maternally inherited allele could be detected (Fig. 2c). These results confirm that the paternally inherited *PRL* allele is not transcribed in either embryo or endosperm during early seed development. Transcripts from both *PRL* alleles can be detected at 156 h.a.p. (Fig. 2c), indicating that the activation of the paternally inherited *PRL* allele occurs only later during seed development. Additional histochemical analysis of ET1041 showed that paternally derived *GUS* expression can be observed earlier, in seeds containing embryos at the mid-globular stage 92 h.a.p. (Fig. 2e). Notably, *GUS* expression is initially restricted to a small region of the chalazal chamber in the nodular cyst of the developing endosperm (Fig. 2d–f), and progressively expands to include cellularizing tissue (data not shown). These results indicate that the paternal silencing that we observed using transposants affects endogenous loci and is independent of the presence of a transgene.

To test whether a particular subset of genes was affected by this epigenetic regulation, we determined the molecular nature of the genes identified by enhancer detection. We isolated genomic regions flanking the *Ds* insertion using thermal asymmetric interlaced PCR (TAIL-PCR)¹⁵. We subcloned and sequenced a total of 16 flanking fragments representing 12 loci (Table 1). For all of them, the insertion site could be confirmed by the presence of a 3' or 5' *Ds* end bordering the genomic sequence. We identified eight insertions within the regulatory region or coding sequence of genes that are represented in GenBank, and two insertions in new genomic sequence. Whereas sequences from four insertions show no homology to sequences in public databases or have similarity to genes of unknown function, the remaining eight genes encode proteins with a wide variety of functions. Some of them are similar to genes that are involved in basic cellular functions such as cell-cycle regulation, the basal transcription machinery or the assembly of protein secondary structure. Others encode putative signal transduction proteins or transcription factors that may have regulatory roles

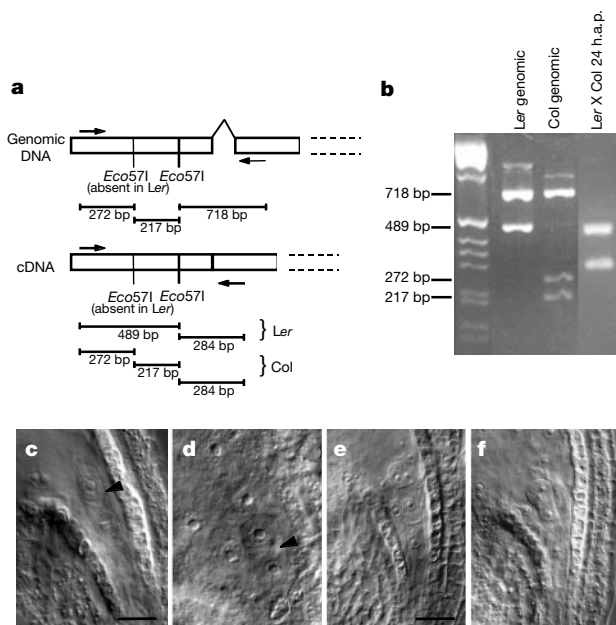


Figure 3 Silencing of the paternally inherited *EMB30* allele during early embryogenesis. **a**, Schematic of *EMB30* showing the position of exons (squared boxes), introns (lines) and the SNP in an *Eco57I* restriction site which is present in *Col* but absent from *Ler*. **b**, Allele-specific RT-PCR analysis of *EMB30*. In reciprocal crosses between *Col* and *Ler* plants, only the transcript derived from the maternal *EMB30* allele is detected 24 h.a.p. **c**, Two-cell wild-type embryo; the first division of the zygote (arrowhead) gives rise to a small apical and an elongated basal cell. **d**, Two-cell *emb30^{tr}/EMB30^p* embryo showing a nearly symmetrical plane of division (arrowhead). **e**, Wild-type embryo with the division plane of the apical cell parallel to the apical-basal axis. **f**, *emb30^{tr}/EMB30^p* embryo with an oblique division plane of the apical cell. Scale bars, 20 μ m (**c,d**); 18 μ m (**e,f**).

during seed development. The *Ds* insertions are located on at least four out of the five chromosomes of *Arabidopsis* (Table 1). These results indicate that the genes identified in our screen do not encode members of a specific family of proteins and are distributed throughout the *Arabidopsis* genome.

Our observations suggest that most, if not all, of the paternal genome is silenced during early seed development. We expect that other genes known to be expressed at early stages are also affected by this regulation; therefore, we examined *EMB30/GNOM*^{16,17} expression using allele-specific RT-PCR. Embryos homozygous for *emb30* are defective in the establishment of the apical-basal axis. In some *emb30* embryos, the zygote divides almost symmetrically, giving rise to an enlarged apical cell that subsequently forms an abnormal globular embryo⁴. Self-pollinated heterozygous *emb30/EMB30* individuals produce 25% of aborted embryo-lethal seeds, suggesting a

zygotic requirement. The phenotype of *emb30* has been interpreted as being caused by a recessive mutation which affects a gene that is active during the earliest diploid (sporophytic) phase of embryogenesis¹⁸. The *emb30* embryo phenotype has been the strongest argument in favour of early genome activation in *Arabidopsis*. *EMB30* encodes a Sec-7-like protein that has been shown to be essential for auxin transport¹⁹. Although *EMB30* is expressed throughout the plant, its allele-specific pattern of expression has not been investigated during seed development. An SNP in the first exon of *EMB30* creates an *Eco57I* site present in *Col* but not in *Ler* (Fig. 3a). We examined the allelic pattern of *EMB30* expression using RT-PCR followed by endonuclease digestion with *Eco57I*. At 24 h.a.p., *EMB30* transcripts from the paternal allele could not be detected (Fig. 3b), suggesting that the initial post-fertilization expression of *EMB30* is exclusively dependent on transcription from the maternally inherited allele. Because *EMB30* may be expressed in sporophytic tissues such as the silique or seed coat, however, we could not discard the possibility that transcripts from the *EMB30* maternal allele are present in vast excess over the paternally derived transcripts, obscuring their detection. Therefore, we tested for genetic activity of the paternal *EMB30* allele by crossing heterozygous *emb30/EMB30* plants with wild-type pollen. If the paternally inherited *EMB30* allele is active after fertilization, heterozygous *emb30^{tr}/EMB30^p* embryos should develop normally. Owing to low expressivity of the *emb30* phenotype at young stages, only 12.2% ($n = 131$) of the embryos derived from self-fertilization of *emb30/EMB30* plants show morphological defects 24 to 48 h.a.p. If the same plants were crossed to wild-type pollen, 12.9% ($n = 116$) of the F_1 seeds showed the same morphological defects, which strongly suggests that the paternally inherited wild-type allele does not provide *EMB30* activity. No defective F_1 embryos were observed when female wild-type plants were crossed with the same heterozygous *emb30/EMB30* individuals ($n = 53$). The defects in *emb30^{tr}/EMB30^p* embryos include a nearly symmetrical plane of division of the zygote (Fig. 3c, d), an oblique plane of division of the apical cell (Fig. 3e, f) and delayed differentiation of the epidermal precursor cells (data not shown). All these abnormalities have previously been described in *emb30* embryos⁴. At maturity, all F_1 seeds resulting from *emb30/EMB30* plants crossed to the wild type are morphologically normal and viable ($n = 157$). Thus, late expression of a paternally derived *EMB30* allele is sufficient to rescue embryos inheriting a maternal *emb30* allele, or, in genetic terms, *emb30* is a paternally rescuable maternal-effect mutant and not a purely zygotic embryo lethal.

Our results indicate that the activity of many genes acting during early embryo and endosperm formation may depend solely on transcription from the maternally inherited allele. These maternal transcripts probably represent a combination of mRNAs transcribed before and after fertilization. The identification of enhancer detector lines in which *GUS* expression is detected in the central cell

Table 1 Identification of genes expressed during early seed formation in *Arabidopsis*

Line	Post-fertilization <i>GUS</i> expression*	Localization of the insertion	Chromosome position†	Accession number‡
ET1041	Embryo and endosperm	PROLIFERA ¹⁴	IV, 12 cM	L39954
ET1051	Embryo and endosperm	Homology to rape mRNA	IV, 69 cM	AL031326 (CAA20464)
ET1275	Embryo	Cop1-interacting protein CIP8 ²⁹	V, 125 cM	AF162150
ET1278	Embryo	Isoflavone reductase homologue		New sequence
ET1811	Embryo and endosperm	14-3-3 GF14 mu	II, 83 cM	AC007087 (AAD 23005)
ET2209	Endosperm	Transmembrane protein homologue	II, 79 cM	AC004261 (AAD12006)
ET2612	Embryo and endosperm	Basal pol III transcription factor homologue	II, 1 cM	AC006200 (AAD14528)
ET2567	Endosperm	Hypothetical protein	IV, 77 cM	AL021749 (CAA16882)
ET3536	Endosperm	AT-hook DNA-binding protein	IV, 69 cM	AL021635 (CAA16562)
ET3988	Embryo and endosperm	Disulphide isomerase homologue	II, 87 cM	AC002535 (AAC62863)
ET3992	Embryo and endosperm	No homology	I, 70 cM	New sequence
ET4336	Embryo and endosperm	Germinating seed mRNA		AF162845

* All lines show post-fertilization *GUS* expression only when self-pollinated or used as females in reciprocal crosses to the wild type.

† The map position is based on the nearest genetic marker referenced in the recombinant inbred list of Lister and Dean (see <http://genome-www.stanford.edu/Arabidopsis/>).

‡ Accession numbers correspond to nucleotide sequences in GenBank; protein accession numbers are indicated in brackets.

only after fertilization or in specific cells of the embryo (data not shown) indicates that at least some of the detected genes are transcribed zygotically from the maternal allele only, and thus are regulated by genomic imprinting as has been shown for the *medea* locus, which displays a true maternal effect^{20,21}. Together, these results strongly suggest that early seed formation in *Arabidopsis* is characterized by a delayed transcriptional activation of the paternally inherited genome. Reporter gene expression and RT-PCR analyses indicate that transcription of paternally inherited alleles initiates late after fertilization, when the embryo consists of 32–64 cells.

Although a case of global paternal silencing has not been reported in the plant kingdom, it is reminiscent of genome-wide heterochromatinization in scale insects^{22,23} and parent-specific X-chromosome inactivation in the extra-embryonic tissues of mammals²⁴. The silencing of the paternal genome probably occurs during sperm cell differentiation and may be related to the tight packaging of sperm chromatin involving specific histones^{25,26}, or changes in methylation levels of sperm DNA as compared with DNA in the vegetative nucleus²⁷. Whatever the mechanism, paternal silencing prolongs the functionally haploid phase, which serves to eliminate deleterious mutations², leading to a more stringent selection against such mutations inherited from the mother. This mechanism is imperfect, however, because only a subset of genes are expressed in the female gametophyte and early seed, and some maternal defects can be paternally rescued as illustrated by *emb30*. The genetic analysis of *emb30* indicates that some of the early embryo-lethal mutants, which have been interpreted as affecting zygotically transcribed genes, represent loci that are only transcribed from the maternal allele. The delayed transcription of a paternally inherited wild-type allele is sufficient to rescue heterozygous embryos carrying a maternally inherited mutant allele, resulting in the 25% aborted seeds observed for many mutants that act during the pre-globular stage of embryogenesis. Thus, paternal silencing is not necessarily reflected in the segregation ratio of aborted seeds in mature siliques. Our results strongly suggest that the first few days of embryogenesis and endosperm development are mostly, if not exclusively, under maternal control. This finding should lead to a reinterpretation of the genetic basis and molecular mechanisms that regulate early seed development in plants. □

Methods

Plant material and growth conditions

Plant growth conditions and insertional mutagenesis have been described^{13,28}. The wild-type strains used were *Arabidopsis thaliana* (L.) Heynh. var. Landsberg (*erecta* mutant: Ler) and *A. thaliana* (L.) Heynh. var. Columbia (Col.). Seeds of *emb30-3/EMB30-3* plants were obtained from the Arabidopsis Biological Resource Center (CS6322).

GUS Assays

Developing carpels and siliques were dissected to expose the ovules and incubated in GUS staining buffer (10 mM EDTA, 0.1% Triton X-100, 2 mM Fe²⁺CN, 2 mM Fe³⁺CN, 100 µg ml⁻¹ chloramphenicol, 1 mg ml⁻¹ X-Gluc (Biosynth) in 50 mM sodium phosphate buffer pH 7.0) for 3 days at 37 °C. The tissue was cleared in 20% lactic acid/20% glycerol and observed on a Leica DMBR microscope under Nomarski optics.

TAIL-PCR

Genomic fragments flanking *Ds* insertions were isolated by TAIL-PCR as described^{15,20} and sequenced at the IOWA State University DNA Sequencing and Synthesis Facility after subcloning into pCRII-TOPO (Invitrogen).

In-situ hybridization

Subcloned TAIL-PCR fragments were linearized with restriction enzymes cutting in the polylinker (*Xho*I, *Bam*HI). Probe synthesis using 1 µg as template and *in situ* hybridization were done as described²¹.

RT-PCR

For RNA preparation, young siliques were harvested in liquid nitrogen at specific time points after pollination. RNA preparation, complementary DNA synthesis and PCR amplification of one-third of the cDNA were done as described^{20,21}. For PRL, the primers used were PRLS1 (5'-CAGTCACTGGTTCATTCCT-3') and PRLAS1 (5'-GTAA-

CAACTCGTCAACAGC-3'). For *EMB30*, the primers used were EMB30S2 (5'-CGCCTAAAGTTGCAATTCCTG-3') and EMB30AS3 (5'-AATCACTGTCTACTCCAGC-3'). PCR products were digested with *Xho*I (PRL) and *Eco*57I (*EMB30*) overnight at 37 °C.

Histological analysis

Siliques were dissected with hypodermic needles (Becton-Dickinson, 1 ml insulin syringes) and fixed in FAA (4% formaldehyde, 5% acetic acid, 50% ethanol). Dissected siliques or individual seeds were cleared in Herr's solution (2:2:2:1 lactic acid:chloral hydrate:phenol:clove oil:xylene, by volume) and observed on a Leica DMRB microscope under brightfield or Nomarski optics.

Received 17 November 1999; accepted 4 January 2000.

- Mordhorst, A. P., Toonen, M. A. J. & de Vries, S. C. Plant embryogenesis. *Crit. Rev. Plant Sci.* **16**, 535–576 (1997).
- Walbot, V. Sources and consequences of phenotypic and genotypic plasticity in flowering plants. *Trends Plant Sci.* **1**, 27–32 (1996).
- Meinke, D. W. Embryo-lethal mutants of *Arabidopsis thaliana*: analysis of mutants with a wide range of lethal phases. *Theor. Appl. Genet.* **69**, 543–552 (1985).
- Mayer, U., Büttner, G. & Jürgens G. Apical-basal pattern formation in the *Arabidopsis* embryo: studies on the role of the *gnom* gene. *Development* **117**, 149–162 (1993).
- van Went, J. L. & Willems, M. T. M. in *Embryology of Angiosperms* (ed. Johri, B.) 273–318 (Springer, Berlin, 1984).
- Bowman, J. L. & Mansfield, S. G. in *Arabidopsis: An Atlas of Morphology and Development* (ed. Bowman, J.) 351–361 (Springer, New York, 1994).
- Berger, F. Endosperm development. *Curr. Opin. Plant Biol.* **2**, 28–32 (1999).
- Zalokar, M. Autoradiographic study of protein and RNA formation during early development in *Drosophila* eggs. *Dev. Biol.* **49**, 425–437 (1976).
- Newman, E. D. & Rothman, J. H. The maternal-to-zygotic transition in embryonic patterning of *Caenorhabditis elegans*. *Curr. Opin. Genet. Dev.* **8**, 472–480 (1998).
- Bensaude, O., Babinet, C., Morange, M. & Jacob, F. Heat shock proteins, first major products of zygotic gene activity in mouse embryo. *Nature* **305**, 331–332 (1983).
- Zimmerman, J. L. & Cohill, P. R. Heat shock and thermotolerance in plant and animal embryogenesis. *New. Biol.* **7**, 641–650 (1991).
- Meinke, D. W. in *Arabidopsis* (eds Meyerowitz, E. M. & Somerville, C. R.) 253–295 (Cold Spring Harbor Lab. Press, Cold Spring Harbor, 1994).
- Sundaresan, V. et al. Patterns of gene action in plant development revealed by enhancer trap and gene trap transposable elements. *Genes Dev.* **9**, 1797–1810 (1995).
- Springer, P. S., McCombie, W. R., Sundaresan, V. & Martienssen, R. A. Gene trap tagging of *PROLIFERA*, an essential MCM2-3-5-like gene in *Arabidopsis*. *Science* **268**, 877–880 (1995).
- Liu, Y.-G., Mitsukawa, N., Oosumi, T. & Whittier, R. F. Efficient isolation and mapping of *Arabidopsis thaliana* T-DNA insert junctions by thermal asymmetric intercalated PCR. *Plant J.* **8**, 457–463 (1995).
- Shevell, D. E. et al. *EMB30* is essential for normal cell division, cell expansion, and cell adhesion in *Arabidopsis* and encodes a protein that has similarity to Sec7. *Cell* **77**, 1051–1062 (1994).
- Busch, M., Mayer, U. & Jürgens, G. Molecular analysis of the *Arabidopsis* pattern formation of gene *GNOM*: gene structure and intragenic complementation. *Mol. Gen. Genet.* **250**, 681–691 (1996).
- Howell, S. H. *Molecular Genetics of Plant Development* (Cambridge Univ. Press, New York, 1999).
- Steinmann, T. et al. Coordinated polar localization of auxin efflux carrier *PIN1* by *GNOM* ARF GEF. *Science* **286**, 316–318 (1999).
- Grossniklaus, U., Vielle-Calzada, J.-P., Hoepfner, M. A. & Gagliano, W. B. Maternal control of embryogenesis by *MEDEA*, a *Polycomb* group gene in *Arabidopsis*. *Science* **280**, 446–450 (1998).
- Vielle-Calzada, J.-P. et al. Maintenance of genomic imprinting at the *Arabidopsis medea* locus requires zygotic DDM1 activity. *Genes Dev.* **13**, 2971–2982 (1999).
- Nur, U. Heterochromatinization and euchromatinization of whole genomes in scale insects (Coccoidea: Homoptera). *Development* (Suppl.) **1**, 29–34 (1990).
- Buglia, G., Predazzi, V. & Ferraro, M. Cytosine methylation is not involved in the heterochromatinization of the paternal genome of mealybug *Planococcus citri*. *Chrom. Res.* **7**, 71–73 (1999).
- Goto, T. & Monk, M. Regulation of X-chromosome inactivation in development in mice and humans. *Microbiol. Mol. Biol. Rev.* **62**, 362–368 (1998).
- Ueda, K. & Tanaka, I. The appearance of male gamete-specific histones gH2B and gH3 during pollen development in *Lilium longiflorum*. *Dev. Biol.* **169**, 210–217 (1995).
- Xu, H., Swoboda, I., Bhalla, P. L. & Singh, M. B. Male gametic cell-specific expression of H2A and H3 histone genes. *Plant Mol. Biol.* **39**, 607–614 (1999).
- Oakeley, E. J., Podesta, A. & Jost, J. P. Developmental changes in DNA methylation of the two tobacco pollen nuclei during maturation. *Proc. Natl Acad. Sci. USA* **94**, 11721–11725 (1997).
- Moore, J. M., Vielle-Calzada, J.-P., Gagliano, W. B. & Grossniklaus, U. Genetic characterization of *hadad*, a mutant disrupting female gametogenesis in *Arabidopsis thaliana*. *Cold Spring Harbor Symp. Quant. Biol.* **62**, 35–47 (1997).
- Torii, K. U. et al. The RING finger motif of photomorphogenic repressor COP1 specifically interacts with the RING-H2 motif of a novel *Arabidopsis* protein. *J. Biol. Chem.* **274**, 27674–27681 (1999).

Acknowledgements

We thank J. Moore and W. Gagliano for help in generating transposants, J. Thomas, A. Coluccio and D. Page for technical assistance, E. Vollbrecht and M. A. Follmer for reviewing the manuscript, and R. Pruitt, V. Sundaresan and E. Vollbrecht for continuous interest in this project. This work was supported in part by the Cold Spring Harbor Laboratory President's Council, a grant from the NRICG Program of the US Department of Agriculture to U.G., a fellowship of the Fonds National Suisse de la Recherche Scientifique to J.-P.V.-C., and scholarships of the Janggen Poehn-Foundation and the Searle Family Trust to U.G.

Correspondence and requests for materials should be addressed to U.G. (grossnik@fmi.ch). New sequences (Table 1) have been submitted to GenBank, (accession numbers AF 228755, AF 228757).

Eomesodermin is required for mouse trophoblast development and mesoderm formation

Andreas P. Russ^{*†‡}, Sigrid Wattler^{§||}, William H. Colledge[†], Samuel A. J. R. Aparicio^{*‡§}, Mark B. L. Carlton^{*†‡}, Jonathan J. Pearce^{*}, Sheila C. Barton^{*}, M. Azim Surani^{*}, Kenneth Ryan^{*||}, Michael C. Nehls^{§||}, Valerie Wilson[#] & Martin J. Evans^{*||}

^{*} Wellcome/CRC Institute for Cancer and Developmental Biology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QR, UK

[†] Department of Physiology, University of Cambridge, Downing Street, Cambridge CB2 3EG, UK

[‡] Paradigm Therapeutics Ltd, Downing Street, Cambridge CB2 3EG, UK

[§] Niedersächsisches Institut für Peptidforschung, Feodor-Lynen-Strasse 31, 30625 Hannover, Germany

^{||} Dep. of Oncology, Cambridge Institute of Medical Research, Addenbrookes' Hospital, Cambridge CB2 2XY, UK

[#] Centre of Genome Research, University of Edinburgh, Kings Buildings, West Mains Road, Edinburgh EH9 3JQ, UK

The earliest cell fate decision in the mammalian embryo separates the extra-embryonic trophoblast lineage, which forms the fetal portion of the placenta, from the embryonic cell lineages. The body plan of the embryo proper is established only later at gastrulation, when the pluripotent epiblast gives rise to the germ layers ectoderm, mesoderm and endoderm. Here we show that the T-box gene *Eomesodermin*¹ performs essential functions in both trophoblast development and gastrulation. Mouse embryos lacking *Eomesodermin* arrest at the blastocyst stage. Mutant trophoblasts do not differentiate into trophoblast, indicating that *Eomesodermin* may be required for the development of trophoblast stem cells². In the embryo proper, *Eomesodermin* is essential for mesoderm formation. Although the specification of the anterior–posterior axis and the initial response to mesoderm-inducing signals is intact in mutant epiblasts, the prospective mesodermal cells are not recruited into the primitive streak. Our results indicate that *Eomesodermin* defines a conserved molecular pathway controlling the morphogenetic movements of germ layer formation and has acquired a new function in mammals in the differentiation of trophoblast.

The T-box genes encode a family of transcription factors sharing an evolutionarily conserved DNA-binding domain first defined in the *Brachyury* (*T*) gene^{3–5}. The T-box gene *Eomesodermin* has been implicated as an important regulator of gastrulation in *Xenopus*¹. To investigate the role of *Eomesodermin* in mammalian development, we mutated the murine orthologue by homologous recombination.

Expression of mouse *Eomesodermin* (*Eomes*) is first detected in the trophoblast lineage, starting in the trophoblast of the blastocyst (Fig. 1a), and continuing in the extra-embryonic ectoderm of the early postimplantation embryo (Fig. 1b–h). In the embryo proper, expression of *Eomes* starts in the posterior part of the epiblast and then extends distally into the primitive streak and nascent mesoderm (Fig. 1c–h). *Eomes* is also expressed in the developing forebrain and the olfactory lobes (Fig. 1i, j)^{6–8}.

We disrupted *Eomes* by inserting a *LacZ* reporter gene into the locus (Fig. 2a–c). In heterozygous embryos, this allele (*Eo*^{LacZ}) shows the same expression pattern as seen by *in situ* hybridization,

albeit with a slight lag in developmental stage most probably reflecting differences between protein and messenger RNA half-lives (Fig. 1e–h). Although animals heterozygous for *Eo*^{LacZ} are healthy and fertile, no homozygous offspring were obtained, indicating embryonic lethality. Morphological analysis at 6.0 and 7.5 days post coitum (d.p.c.) revealed that mutant embryos arrest soon after implantation and fail to form organized embryonic or extra-embryonic structures (Fig. 3a–e). This impairment of peri-implantation development is probably due to a defect in the trophoblast lineage, as *Eomes* is only expressed in the trophoblast at this stage.

The differentiation of mutant trophoblast was further investigated by blastocyst culture *in vitro*. In control embryos (*n* = 99), trophoblast cells started to spread on the culture dish 15–24 h after hatching and usually formed extensive outgrowths by 48 h (Fig. 3f). In contrast, most homozygotes (30 out of 33) did not form outgrowths, although they hatched normally and maintained the typical blastocyst morphology for more than 7 days (Fig. 3g). Some mutants (3 out of 33) showed late (>96 h) spreading of small trophoblast-like cells, which then rapidly degenerated (not shown). Mutant blastocysts expressed the typical repertoire of integrins (Fig. 3h–k) and attached strongly to the substratum, indicating that the defect in trophoblast differentiation is not due to a lack of integrin-mediated adhesion^{9,10}.

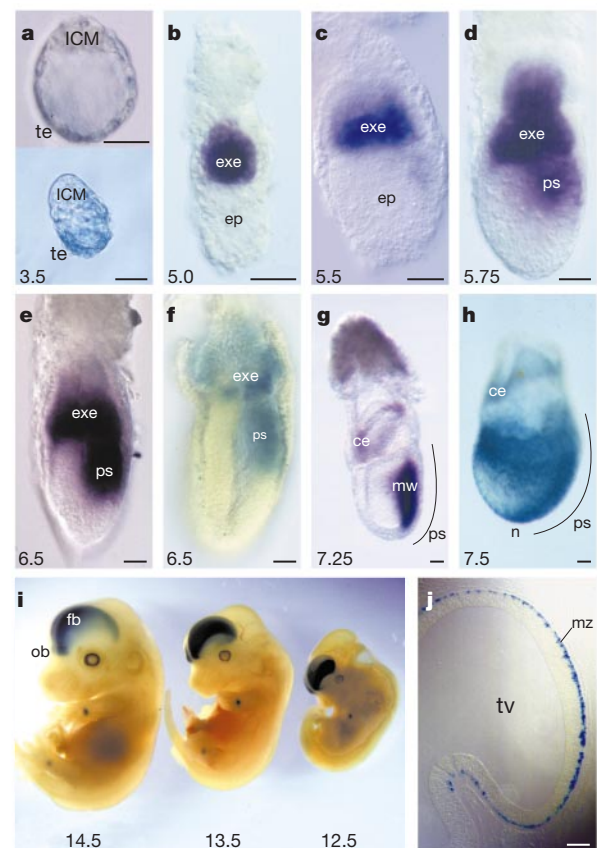


Figure 1 Expression of *Eomesodermin* from 3.5 to 14.5 d.p.c. as shown by β -galactosidase staining of *Eo*^{LacZ} heterozygotes (a, f, h–j) or whole-mount *in situ* hybridization (b–e, g), anterior aspects facing left. a–f, Expression in the trophoblast (a), extra-embryonic ectoderm and posterior epiblast (b–f). g, h, Expression in primitive streak, mesoderm wings and chorionic ectoderm. i, Expression in forebrain and limbs. j, Coronal section of forebrain at 12.5 d.p.c. te, trophoblast; ICM, inner cell mass; exe, extra-embryonic ectoderm; ep, epiblast; ps, primitive streak; mw, mesoderm wings; ce, chorionic ectoderm; n, node; fb, forebrain; ob, olfactory bulb; mz, mantle zone; tv, telencephalic vesicle. Scale bars, 50 μ m.

|| Present addresses: Ingenium Pharmaceuticals AG, Lochhamer Strasse 29, 82152 Martinsried, Germany (S.W.; M.C.N.); Children's Hospital of Philadelphia, University of Pennsylvania School of Medicine, 34th and Civic Center Blvd, Philadelphia, PA 19104, USA (K.R.); Cardiff School of Biosciences, Museum Avenue, PO Box 911, Cardiff CF10 3US, Wales, UK (M.J.E.).

A self-renewing population of diploid trophoblast cells can be derived *in vitro* from blastocyst outgrowths². These trophoblast stem (TS) cell lines strongly express *Eomes*, and their maintenance requires the presence of fibroblast growth factor 4 (FGF4). When blastocysts lacking *Eomes* were cultured in the presence of FGF4, the block in trophoblast differentiation could not be overcome, indicating that *Eomes* may be required cell autonomously for the transition from trophoblast to trophoblast.

We investigated the role of *Eomes* in gastrulation using chimaeras consisting of wild-type extra-embryonic and mutant embryonic tissues. In tetraploid host embryos, which can only generate extra-embryonic tissues, the phenotype of mutant epiblasts derived from homozygous embryonic stem (ES) cells or inner cell masses (ICMs) can be studied¹¹.

No gross abnormalities were observed in chimaeras at early gastrulation stages (6.5 d.p.c., Fig. 4a). At 7.5 d.p.c., the typical elongation of the primitive streak and a morphologically distinct node were absent (Fig. 4b). Histological sections show a posterior thickening of the epiblast with morphological signs of epithelial-to-mesenchymal transition, but no emergence of embryonic (Fig. 4e) or extra-embryonic mesoderm (Fig. 4f, g). At 8.5 d.p.c., the equivalent of early somite stages, mutant chimaeras still resembled enlarged egg cylinders lacking mesoderm (Fig. 4c, d).

This phenotype could be due to defects in axis specification and mesoderm induction, or due to an impairment in the formation or function of the primitive streak. To distinguish between these possibilities, we analysed the expression of marker genes. In mutant epiblasts, the targeted *Eo^{LacZ}* locus is initially expressed in its normal posterior domain (6.5 d.p.c., Fig. 4a), but subsequently shows broad ectopic activation in the lateral and anterior parts of

the embryonic ectoderm (Fig. 4b–d). The expression of *Brachyury/T* and *Fgf8*, two early markers of the primitive streak and prospective mesoderm, is similarly extended (compare Fig. 4b and 4h, i). This indicates that the anterior–posterior axis is specified and mesoderm inducing signals are present. The domain of ectopic marker expression is similar to the location of prospective mesoderm in the pregastrulation epiblast¹², suggesting that prospective mesodermal cells receive an inducing signal and activate early streak genes, but are unable to initiate the directed migration towards the streak. No expression of *Hnf-3β* is detected in mutants (data not shown), indicating that the differentiation of anterior streak and node is blocked.

Ectodermal patterning was visualized by *Otx2*, which is normally active in undifferentiated epiblast cells and neuroectoderm, and *Hesx1*, a marker of anterior visceral endoderm and neuroectoderm¹³. In mutant epiblasts, *Otx2* is expressed only in the anterior domain (Fig. 4j), indicating that its typical suppression by posterior signals is intact. No ectodermal expression of *Hesx1* was detected (Fig. 4l), suggesting that the mutant anterior ectoderm is undifferentiated, and that the specification of neuroectoderm is delayed or absent.

The anterior visceral endoderm (AVE) has been implicated as the main signalling centre for the anterior patterning of the mouse embryo¹⁴. Expression of *Hesx1* is reproducibly seen in the AVE, which is of wild-type origin in chimaeras (Fig. 4l), indicating that the specification and positioning of the AVE is not impaired in the presence of a mutant epiblast.

To test whether *Eomes* is required for mesoderm differentiation in general, or specifically for the recruitment of prospective mesoderm into the primitive streak, we examined the developmental potential

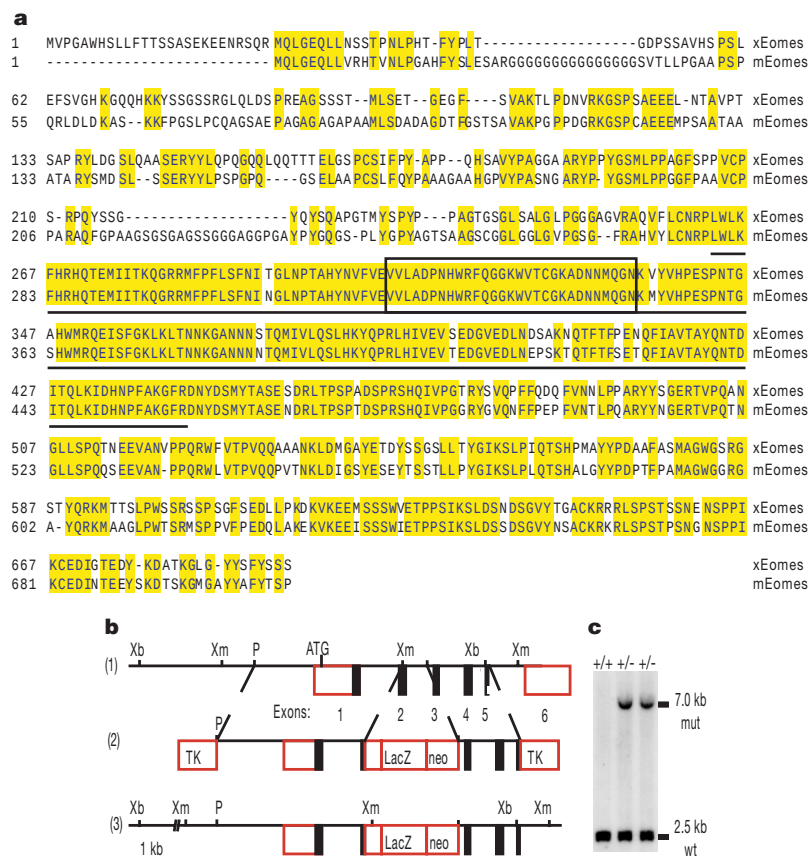


Figure 2 Primary structure and targeted disruption of mouse *Eomesodermin*. **a**, Alignment of the predicted amino-acid sequences of mouse (*mEomes*) and *Xenopus* (*xEomes*) proteins. Identical residues are shaded, the T domain is underlined. The region deleted in the *Eo^{LacZ}* allele is boxed. **b**, Targeted disruption of the *Eomesodermin* locus: wild-type locus (1), targeting construct (2), mutant allele (3). The T-box coding region is shaded. Xb, *Xba*I; Xm, *Xmn*I, P, *Pst*I. **c**, Southern blot with 3' external probe, *Xmn*I digest.

the *Eo^{LacZ}* allele is boxed. **b**, Targeted disruption of the *Eomesodermin* locus: wild-type locus (1), targeting construct (2), mutant allele (3). The T-box coding region is shaded. Xb, *Xba*I; Xm, *Xmn*I, P, *Pst*I. **c**, Southern blot with 3' external probe, *Xmn*I digest.

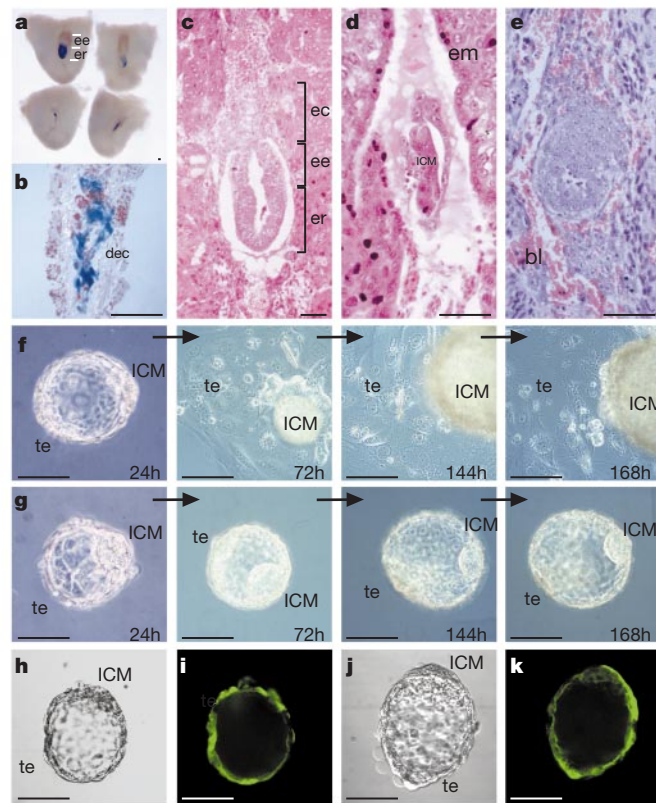


Figure 3 Morphology of mutant embryos. **a**, Heterozygous (top) and homozygous embryos (bottom) in decidua, LacZ stain. **b**, Longitudinal section through disorganized homozygous embryo in decidua, LacZ stain. **c–e**, Haematoxylin and eosin (HE)-stained longitudinal sections of embryos in decidua at 6.5 d.p.c. **c**, Wild-type littermate. **d**, Presumptive homozygous mutant arrested in a blastocyst-like stage. **e**, Disorganized embryo in a blood-filled implantation site (bl). **d, e** are depicted at higher magnification than **c, f, g**. **f–k**, Time Course of mutant blastocysts and control littermates cultured *in vitro*.

Phase contrast micrographs. **f**, Control embryo. Trophoblast-derived cells (te) are spreading over the dish, whereas ICM-derived cells form a central clump. **g**, Typical mutant blastocyst showing no sign of trophoblast outgrowth. **h–k**, Expression of $\alpha 5/\beta 1$ integrin in mutant blastocyst (**h, i**) and wild-type littermate (**j, k**) 36 h after hatching. Phase contrast (**h, j**); immunofluorescence (**i, k**). em, endometrium; er, embryonic region; ee, extra-embryonic ectoderm; dec, decidua; ec, ectoplacental cone; ICM, inner cell mass; te, trophoblast and derivatives. Scalebars, 50 μ m.

of homozygous mutant ES cells. Mesoderm formation is partially rescued in chimaeras with diploid host embryos, where mutant and wild-type epiblast cells are intermingling (Fig. 4m, n). Wild-type cells migrate efficiently through the primitive streak and become enriched in the mesoderm wings, whereas most mutant cells persist in the ectoderm (Fig. 4n). When injected into syngeneic hosts, mutant cells give rise to teratomas containing muscle, cartilage and haematopoietic tissue (Fig. 4o). This shows that *Eomes* is not absolutely required for the differentiation of mesodermal cell types, and indicates that the mutation may specifically affect the morphogenetic movement of cells from the epiblast into the primitive streak.

In *Xenopus*, the homeobox-containing genes *Bix1–4* have been identified as transcriptional targets for the T-box genes *Brachyury/Xbra* and *VegT* (refs 15, 16). A murine homologue of the *Bix/Mix* gene family is *Mml*, which is normally expressed in the visceral endoderm and the primitive streak¹⁷. To test whether *Mml* is a potential transcriptional target of T-box genes in the mouse, we analysed its expression in *Eomes*- and *Brachyury*-deficient (*T/T*) embryos. *Mml* expression is normal in *T* mutants (not shown). No transcription could be detected in epiblasts lacking *Eomes* (Fig. 4k), whereas the other early streak markers tested show enhanced expression (see above). This suggests that *Mml* is acting downstream of *Eomes*.

Our findings show that *Eomes* performs independent essential functions in extra-embryonic and embryonic tissues. The expression of *Eomes* in trophoblast, diploid trophoblast and TS cell lines, and the phenotype of mutant embryos *in vivo* and *in vitro*, indicate that *Eomes* may be required for the differentiation of

trophoblast and the formation of trophoblast stem cells^{2,18}. Notably, although the differentiation of trophoblast into primary giant cells *in vitro* is dependent on *Eomes*, it does not require FGF4, as shown by mutations in the FGF pathway^{19,20}. This suggests that FGF4 is not absolutely required for the initial activation of *Eomes* in trophoblast, but that it might act to maintain *Eomes* expression and to expand the TS cell pool.

In the embryo proper, *Eomes* is required for mesoderm formation and the morphogenetic movements of gastrulation. In contrast to other mutations affecting gastrulation^{21–23}, lack of *Eomes* abrogates the formation of embryonic and extraembryonic mesoderm without apparent effects on the initial specification of the anterior–posterior axis, mesoderm induction, cell viability or proliferation. The spatiotemporal expression patterns of marker genes suggest that *Eomes* is specifically required for the directed movement of cells from the epiblast into the streak in response to mesoderm induction. The accompanying block in mesoderm differentiation is probably secondary to the defect in migration, as mutant cells can give rise to mesodermal tissues *in vitro* and in teratomas.

Previous work has implicated *Brachyury/T* (refs 24, 25) and FGF signalling^{26,27} in the control of cell migration through the primitive streak. Our data indicate that *Eomes* acts at an earlier stage than *T*, *Fgfr1* and *Fgf8*. A potential transcriptional target of *Eomes* is *Mml* (ref. 17), suggesting that, similar to *Xenopus*, a *Mix*-like homeobox protein might be acting downstream of a T-box gene^{15,16}.

It is intriguing that during evolution a conserved protein that controls mesoderm formation has acquired an additional function in the trophoblast—a ‘new’ tissue type, specific to placental mammals. The common denominator of *Eomes* function in

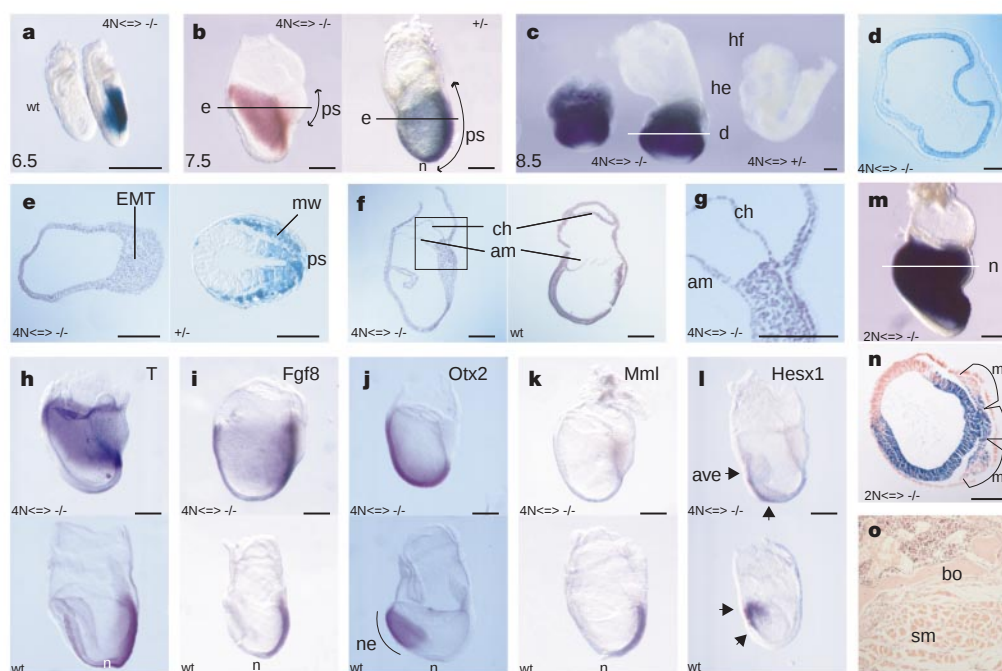


Figure 4 Analysis of chimaeric embryos. **a–l**, Chimaeras derived from homozygous mutant ES cells (**a, b, e–l**) or ICMs (**c, d**) in tetraploid host embryos. **a**, Chimaera (right) and control at 6.5 d.p.c., *LacZ* stain. **b**, Chimaera (left) and heterozygous control, 7.5 d.p.c. The mutant primitive streak is not elongated, *Eo^{lacZ}* is expressed in the posterior, lateral and anterior epiblast. **c**, Two chimaeras and control littermate (right) at 8.5 d.p.c., *LacZ* stain. hf, headfolds; he, heart. **d**, *e*, Transverse sections as indicated in (**c, b**). Chimaeras lack mesoderm and posterior, lateral and anterior epiblast express *Eo^{lacZ}* (**d**). The control (**e**, right) expresses *Eo^{lacZ}* in the primitive streak and mesoderm wings. **f**, Longitudinal sections, 7.5 d.p.c., HE stain. Posterior thickening of the epiblast and irregular anterior ectodermal fold in the chimaera (left). Amnion (am) and chorion (ch) consist only of ectodermal cell layers, extra-embryonic mesoderm is absent. **g**, Higher magnification of area boxed in **f**. **h–l**, Analysis of marker gene expression by *in situ* hybridization in chimaeras (top) and controls (bottom). **h, i**, *T* (*Brachyury*) (**h**) and *Fgf8* (**i**)

are expressed in the same extended posterior–lateral domain as *Eo^{lacZ}*. Controls show normal expression in the primitive streak. **j**, *Otx2* is repressed in the posterior–lateral epiblast. Expression is confined to neuroectoderm (ne) in control. **k**, *Mml* expression is not detected in the mutant epiblast. **l**, In the chimaera, *Hesx1* expression is detected in the anterior visceral endoderm, but not in the epiblast. **m**, Chimaera with high contribution of mutant ES cells in a diploid host at 7.5 d.p.c., *LacZ* stain. **n**, Transverse section as indicated in **m**. Mesoderm wings are predominantly composed of wild-type cells, whereas mutant cells expressing *Eo^{lacZ}* persist in the embryonic ectoderm. **o**, Teratomas derived from homozygous mutant ES cells show extensive differentiation into mesodermal cell types. HE stain. sm, striated muscle; bo, bone; ps, primitive streak; EMT, epithelial-to-mesenchymal transition; n, node; mw, mesoderm wings. All embryos are shown in lateral view, anterior facing to the left. Scalebars, 200 μ m.

trophoblast differentiation and gastrulation might be the control of migratory and invasive properties. □

Methods

Mutant ES cells and mice

Complementary DNAs and genomic clones were isolated as described²⁸. In the targeting vector, parts of exon 2 and intron 2 were replaced with an IRES-*LacZ*-MCneo cassette (Fig. 2b). Germline transmission and homozygous mutant ES cell lines were obtained from two independent ES cell clones. Preimplantation embryos and blastocyst outgrowths were genotyped by multiplex polymerase chain reaction. Embryo culture, *LacZ* staining, immunofluorescence and *in situ* hybridization were performed according to standard protocols²⁹.

Generation of chimaeric embryos and teratomas

Tetraploid host embryos were generated by electrofusion²⁹. Inner cell masses were isolated by immunosurgery from blastocysts derived from heterozygote intercrosses and injected into tetraploid host blastocysts. Homozygous mutant ES cells were aggregated with tetraploid host morulae. Similar results were obtained with homozygous ES cell lines and ICMs. Teratomas were generated by subcutaneous injection of 2×10^6 heterozygous or homozygous mutant ES cells into syngeneic hosts and dissected for histology after 6–8 weeks.

Received 24 September 1999; accepted January 2000.

- Ryan, K., Garrett, N., Mitchell, A. & Gurdon, J. B. Eomesodermin, a key early gene in *Xenopus* mesoderm differentiation. *Cell* **87**, 989–1000 (1996).
- Tanaka, S., Kunath, T., Hadjantonakis, A. K., Nagy, A. & Rossant, J. Promotion of trophoblast stem cell proliferation by FGF4. *Science* **282**, 2072–2075 (1998).
- Herrmann, B. G., Labeit, S., Poustka, A., King, T. R. & Lehrach, H. Cloning of the *T* gene required in mesoderm formation in the mouse. *Nature* **343**, 617–622 (1990).
- Papioannou, V. E. & Silver, L. M. The *T*-box gene family. *Bioessays* **20**, 9–19 (1998).

- Smith, J. C. T-Box genes: What they do and how they do it. *Trends Genet.* **15**, 154–158 (1999).
- Hancock, S. N., Agulnik, S. I., Silver, L. M. & Papaioannou, V. E. Mapping and expression analysis of the mouse ortholog of *Xenopus* eomesodermin. *Mech. Dev.* **81**, 205–208 (1999).
- Ciruna, B. G. & Rossant, J. Expression of the *T*-box gene eomesodermin during early mouse development. *Mech. Dev.* **81**, 199–203 (1999).
- Bulfone, A. *et al.* Expression pattern of the *Tbr2* (Eomesodermin) gene during mouse and chick brain development. *Mech. Dev.* **84**, 133–138 (1999).
- Sutherland, A. E., Calarco, P. G. & Damsky, C. H. Expression and function of cell surface extracellular matrix receptors in mouse blastocyst attachment and outgrowth. *J. Cell Biol.* **106**, 1331–1348 (1988).
- Sutherland, A. E., Calarco, P. G. & Damsky, C. H. Developmental regulation of integrin expression at the time of implantation in the mouse embryo. *Development* **119**, 1175–1186 (1993).
- Rossant, J. & Spence, A. Chimeras and mosaics in mouse mutant analysis. *Trends Genet.* **14**, 358–363 (1998).
- Lawson, K. A., Meneses, J. J. & Pedersen, R. A. Clonal analysis of epiblast fate during germ layer formation in the mouse embryo. *Development* **113**, 891–911 (1991).
- Thomas, P. & Beddington, R. Anterior primitive endoderm may be responsible for patterning the anterior neural plate in the mouse embryo. *Curr. Biol.* **6**, 1487–1496 (1996).
- Beddington, R. S. & Robertson, E. J. Anterior patterning in mouse. *Trends Genet.* **14**, 277–284 (1998).
- Tada, M., Casey, E. S., Fairclough, L. & Smith, J. C. Bix1, a direct target of *Xenopus* T-box genes, causes formation of ventral mesoderm and endoderm. *Development* **125**, 3997–4006 (1998).
- Casey, E. S. *et al.* Bix4 is activated directly by VegT and mediates endoderm formation in *Xenopus* development. *Development* **126**, 4193–4200 (1999).
- Pearce, J. J. H. & Evans, M. J. *Mml*, a mouse Mix-like gene expressed in the primitive streak. *Mech. Dev.* **87**, 189–192 (1999).
- Rossant, J. & Tamura-Lis, W. Effect of culture conditions on diploid to giant-cell transformation in postimplantation mouse trophoblast. *J. Embryol. Exp. Morphol.* **62**, 217–227 (1981).
- Feldman, B., Poueymirou, W., Papaioannou, V. E., DeChiara, T. M. & Goldfarb, M. Requirement of FGF-4 for postimplantation mouse development. *Science* **267**, 246–249 (1995).
- Arman, E., Haffner-Krausz, R., Chen, Y., Heath, J. K. & Lonai, P. Targeted disruption of fibroblast growth factor (FGF) receptor 2 suggests a role for FGF signaling in pregastrulation mammalian development. *Proc. Natl Acad. Sci. USA* **95**, 5082–5087 (1998).
- Liu, P. *et al.* Requirement for Wnt3 in vertebrate axis formation. *Nature Genet.* **22**, 361–365 (1999).
- Ding, J. *et al.* Crip1 is required for correct orientation of the anterior–posterior axis in the mouse embryo. *Nature* **395**, 702–707 (1998).

23. Tam, P. P. & Behringer, R. R. Mouse gastrulation: the formation of a mammalian body plan. *Mech. Dev.* **68**, 3–25 (1997).
24. Wilson, V., Manson, L., Skarnes, W. C. & Beddington, R. S. The T gene is necessary for normal mesodermal morphogenetic cell movements during gastrulation. *Development* **121**, 877–886 (1995).
25. Wilson, V., Rashbass, P. & Beddington, R. S. Chimeric analysis of T (Brachyury) gene function. *Development* **117**, 1321–1331 (1993).
26. Ciruna, B. G., Schwartz, L., Harpal, K., Yamaguchi, T. P. & Rossant, J. Chimeric analysis of fibroblast growth factor receptor-1 (Fgfr1) function: a role for FGFR1 in morphogenetic movement through the primitive streak. *Development* **124**, 2829–2841 (1997).
27. Sun, X., Meyers, E. N., Lewandoski, M. & Martin, G. R. Targeted disruption of Fgf8 causes failure of cell migration in the gastrulating mouse embryo. *Genes Dev.* **13**, 1834–1846 (1999).
28. Wattler, S., Russ, A., Evans, M. & Nehls, M. A combined analysis of genomic and primary protein structure defines the phylogenetic relationship of new members of the T-box family. *Genomics* **48**, 24–33 (1998).
29. Hogan, B., Beddington, R., Constantini, F. & Lacy, E. *Manipulating the Mouse Embryo* (Cold Spring Harbor Laboratory Press, New York, 1994).

Acknowledgements

We thank J. Gurdon, A. Bulfone, A. Zorn, N. Papalopulu, D. St. Johnston and R. Pedersen for discussion; X. Sun and F. Beck for communicating results before publication; G. Martin, C. Wright and R. Milner for gifts of probes and reagents; F. Wianny and A. Sossick for help with confocal microscopy; J. Wilson for technical assistance; J. Ferguson, P. Whiting and R. Plumridge for animal care. A.P.R. thanks W. Gross for continuing support and advice. This work was funded by grants from the Wellcome Trust. V.W. is supported by a Medical Research Council Career Development Award.

Correspondence and request for materials should be addressed to A.P.R.
(e-mail: apr@mole.bio.cam.ac.uk).

p73-deficient mice have neurological, pheromonal and inflammatory defects but lack spontaneous tumours

Annie Yang*, Nancy Walker†, Roderick Bronson‡, Mourad Kaghad‡, Mariette Oosterwegel§, Jacques Bonnin†, Christine Vagner†, Helene Bonnet†, Pieter Dikkes||, Arlene Sharpe§, Frank McKeon* & Daniel Caput†

* Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115, USA

† Sanofi Recherche, Imnopol B.P. 137, 31676 Labege Cedex, France

‡ U.S. Department of Agriculture, Human Nutrition Research Center on Aging, and Department of Pathology, Tufts University School of Veterinary Medicine, Boston, Massachusetts 02111, USA

§ Immunology Research Division, Departments of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA

|| Department of Neurology, Division of Neuroscience, Children's Hospital, Boston, Massachusetts 02115, USA

p73 (ref. 1) has high homology with the tumour suppressor p53 (refs 2–4), as well as with p63, a gene implicated in the maintenance of epithelial stem cells^{5–7}. Despite the localization of the p73 gene to chromosome 1p36.3, a region of frequent aberration in a wide range of human cancers¹, and the ability of p73 to transactivate p53 target genes¹, it is unclear whether p73 functions as a tumour suppressor. Here we show that mice functionally deficient for all p73 isoforms exhibit profound defects, including hippocampal dysgenesis, hydrocephalus, chronic infections and inflammation, as well as abnormalities in pheromone sensory pathways. In contrast to p53-deficient mice, however, those lacking p73 show no increased susceptibility to spontaneous tumorigenesis. We report the mechanistic basis of the hippocampal dysgenesis and the loss of pheromone responses, and show that new, potentially dominant-negative, p73 variants are the predominant expression products of this gene in developing and adult tissues. Our data suggest that there is a marked divergence in

the physiological functions of the p53 family members, and reveal unique roles for p73 in neurogenesis, sensory pathways and homeostatic control.

Our analysis of murine p73 complementary DNAs revealed transcripts encoding p73 proteins that lack an amino-terminal transactivation domain (Δ N-p73), in addition to those that show the carboxy-terminal diversity reported for human p73 (Fig. 1a)^{1,8}. Like the transcripts encoding Δ N-p63 (ref. 5), the Δ N-p73 messages are derived from an alternative promoter located in intron 3. Δ N-p73 failed to activate transcription from a p53-reporter gene (data not shown) but suppressed the transactivation activity of p73 α by hetero-oligomerization and competition for DNA binding (Fig. 1b, c). *In situ* hybridization experiments on mouse embryos failed to detect p73 β , γ or δ transcripts, whereas p73 α messenger RNA was highly expressed in a number of epithelial and neural structures, including nasal epithelium, the vomeronasal organ, the hippocampus and the hypothalamus (Fig. 1d; and data not shown). Additional probes revealed that Δ N-p73 transcripts appeared to be the predominant p73 gene product in adults (data not shown) and during embryogenesis (Fig. 1e).

Mice with disrupted p73 alleles were produced using homologous recombination in embryonic stem (ES) cells to replace exons 5 and 6, which encode the DNA-binding domain, with the neomycin-resistance (*NEO*^R) gene^{9,10} (Fig. 1f). The offspring of p73 heterozygous mice showed a mendelian distribution of genotypes indicating that p73, like p53, is not required for embryogenesis^{11,12} (Fig. 1g). Polymerase chain reaction with reverse transcriptase (RT-PCR) and sequence analysis of RNA from murine embryonic fibroblasts (MEFs) of known p73 genotypes confirmed the disruption of the p73 gene (Fig. 1g).

p73^{-/-} pups showed a runting phenotype and high rates of mortality (Fig. 2a, b). Most commonly, death followed massive gastrointestinal haemorrhages, although intracranial bleeding was apparent in ~15% of the mortalities. Histological examinations of the gastrointestinal tract of the p73^{-/-} postnatal day 7 (P7) mice revealed numerous abnormalities including erosion marked by a loss of enterocytes and excessive mucosecretions in the duodenum, ileum and cecum (Fig. 2c; and data not shown), which may underlie the wasting syndrome and the intestinal haemorrhaging in these mice.

A striking feature of the p73^{-/-} mice was severe rhinitis and purulent otitis media (Fig. 2d; and data not shown). Massive neutrophil infiltrates of these sites were obvious at the earliest ages examined (P2 pups) and persisted through adulthood, when greater than 80% of p73^{-/-} mice exhibited chronic, bilateral rhinitis, otitis, periorbital oedema and conjunctivitis. Microbiological analysis of affected sites from p73^{-/-} weanlings (P21) revealed the presence of *Escherichia coli*, *Pasteurella aerogenes* and micrococcal species. Despite these indications of inflammation and infection, no obvious deficiencies in lymphoid or granulocyte populations were detected in the p73^{-/-} mice, indicating that there might be defects in other components of the natural immune system. In support of this notion, p73 was highly expressed in the epithelia bordering the sites of infections in P21 mice (Fig. 2e). These epithelia function, in part, as barriers to external pathogens through secretion of mucus, cytokines and antimicrobial factors¹³. Despite the massive sinus inflammation marked by mucoserous secretions, neutrophil infiltration and goblet cell hyperplasia of the surrounding epithelia in the P2 pups (data not shown), these pups also lacked obvious Gram-staining pathogens. Together, these data indicate that the infections seen in the older pups may be, as with cystic fibrosis^{14,15}, the result of inappropriate or hyperactive epithelial responses, which in turn render these sinuses habitable by microorganisms. These observations raise the possibilities that the infections in the p73^{-/-} mice are secondary to constitutive inflammatory signals promoted by the loss of p73, and may be linked to the gastrointestinal phenotype as a generalized pan-mucositis.

The $p73^{-/-}$ mice also show a mild, congenital hydrocephalus that progresses in some animals to a highly morbid condition marked by massive expansion of the lateral ventricles, compression of the cortex and intraventricular bleeding (Fig. 2f). Our analysis revealed a communicating form of hydrocephalus in these mice, indicating that there might be defects in production or reabsorption of cerebral spinal fluid (CSF)¹⁶. Precedence for the overproduction of CSF leading to hydrocephalus is found in mice lacking the transcription factor, E2F-5; these mice also develop a nonobstruc-

tive hydrocephalus¹⁷. Moreover, just as E2F-5 is abundant in the choroid plexus¹⁷, p73 is highly expressed in the epithelial cells of the choroid plexus and in the ependymal cells lining the ventricles (not shown), both of which regulate CSF secretion and reabsorption. These data show that p73 is critical in regulating the homeostasis of CSF dynamics and raise the possibility of p73 interactions with E2F-5 in these processes.

Examination of the brains of the $p73^{-/-}$ mice revealed hippocampal dysgenesis, which is characterized by unusual arrangements of the CA1–CA3 pyramidal cell layer and the dentate gyrus (Fig. 3a, b). The dentate gyrus lacked an infrapyramidal blade and the suprapyramidal blade appeared hypertrophied and extended (Fig. 3a, b), which overall suggest an organization defect rather than a deficit of granular neurons. No major structural abnormalities were observed in other areas of the $p73^{-/-}$ brain, including the neocortex and cerebellum. Although defects in neurogenesis have been described



in several mouse models, particularly the *reeler* and *scrambler* mutants, these phenotypes reflect a generalized defect in neuronal migration affecting lamination in the cortex, hippocampus and cerebellum^{18,19}, and are therefore distinct from the dysgenesis observed in the *p73*^{-/-} mice. It was thus interesting that in mid-gestation embryos *p73* expression closely mirrored the expression patterns of reelin, a secreted glycoprotein which is essential for neuronal migration in the cortex^{20–24}. In E12.5 animals, for example, cells lining the outer preplate layer of the telencephalic vesicles, the bed nucleus of the stria terminalis and the preoptic area (Fig. 3c) all stain positively for both reelin and *p73*. The correlation between *p73* and reelin diverged in the wild-type, postnatal (P3) brain, where the two proteins were found to be co-expressed only in a distinct

population of large, bipolar cells, known as the Cajal–Retzius neurons (CR), distributed in a continuous band along the marginal zone of the cortex extending to the molecular layer of the dentate gyrus (Fig. 3d, e). Other sites of reelin expression, including the cerebellum, cortical layer five and the olfactory bulb, were devoid of *p73* signal (Fig. 4a). Notably, the *p73*^{-/-} P3 brain showed an absence of reelin expression specifically in the marginal zone and the molecular layer of the dentate gyrus (Fig. 4b, c), whereas reelin expression in the cerebellum, cortical layer five and the olfactory bulb was similar to that of wild-type mice. In addition to the lack of reelin, we also failed to detect a second CR marker, calretinin, in the cortical and hippocampal marginal zones (data not shown).

These data indicate that functional loss of *p73* leads to the disappearance of the CR neurons in the cortical marginal zone and hippocampal molecular layer, and that this defect underlies the hippocampal dysgenesis seen in *p73*-deficient mice. Notably, the loss of CR neurons in the *p73*^{-/-} mouse is not accompanied by a cortical lamination defect, as seen in the *reeler* mouse. Moreover, the hippocampal dysgenesis of *p73*^{-/-} mice is clearly distinct from the hippocampal defect of the *reeler* mouse¹⁸. Together, these findings indicate that *p73*-dependent CR neurons produce factors other than reelin to effect hippocampal neurogenesis. The link between the absence of these CR neurons and the striking yet highly site-specific dysgenesis in the *p73*^{-/-} brain further suggests that CR

cells, widely believed to control cortical neurogenesis, may instead have a critical role in hippocampal development. Finally, p73 expression is maintained in P21 CR neurons of the hippocampal molecular layer (Fig. 4d), indicating that p73 may function in the modelling of persistent neurogenesis of the dentate gyrus throughout adulthood^{25,26}.

The p73-deficient male mice lacked both interest in sexually mature females and aggressive responses to other males. To quantify this defect, 15 *p73*^{-/-} males, age 6–10 weeks, were housed with three wild-type females each and pregnancies scored after 30 days. Only one pregnancy was obtained from these matings, whereas most females (>90%) housed with wild-type males were pregnant by 30 days. Likewise, no pregnancies were observed in female *p73*^{-/-} mice that mated with wild-type males, indicating a defect in conceiving or maintaining embryos. No structural abnormalities in the reproductive organs of either male or female *p73*^{-/-} mice were observed by histology, suggesting, instead, that they might have defects in hormonal or sensory pathways^{27,28}. The highest levels of p73 RNA in the mouse were found in the neuroepithelium of both the embryonic and adult vomeronasal organ (VNO), an accessory olfactory structure involved in pheromone detection (Fig. 5a). Given this, and the behavioural and reproductive abnormalities in *p73*^{-/-} mice, p73 deficiency might affect VNO function. We examined the expression patterns of the two structurally distinct classes of pheromone receptors present in the VNO, V1R and V2R^{29,30}. The wild-type VNO showed numerous V2R- and V1R-expressing cells, but no cells were positive in the *p73*^{-/-} VNO (Fig. 5b). Similarly, expression of olfactory cell adhesion molecule (OCAM), a marker of neurosensory cells of the main olfactory epithelium (MOE) and

the VNO, was absent from the VNO of *p73*^{-/-} mice (Fig. 5c). OCAM expression was normal in the MOE of the *p73*^{-/-} mouse (data not shown), suggesting a site-specific consequence of p73 loss on these related cell populations. The lack of OCAM and V1R and V2R receptors was not attributable to a gross absence of cells in the VNO, as probes for the olfactory marker protein (OMP), expressed by differentiated chemosensory neurons of the VNO and MOE, revealed strong expression in both wild-type and mutant VNOs (Fig. 5d). Finally, the p73 signal was also intense in secondary and tertiary projections of the accessory olfactory system, including the accessory olfactory bulb, the amygdala and hypothalamus (data not shown), suggesting that additional defects in sensory or hormonal pathways may contribute to the reproductive and behavioural phenotypes of p73-deficient mice.

During this work, we carried out necropsies on over 100 *p73*^{-/-} mice ranging in ages from 2 to 15 months, and found no increased tendency in these mice to develop spontaneous tumours. These data appear to preclude the notion that p73 functions in a manner similar to p53 to suppress tumour development; however, whether Δ N-p73 variants, which potentially counteract p53, modulate p53 activity in the cell remains unresolved.

Our data reveal a marked divergence in the respective developmental activities of p63 and p73 and further distinguish these genes from p53, the most recently evolved member of the family apparently devoted to tumour suppression. Despite these revelations, we know relatively little about specific gene targets for any member of the p53 family of transcriptional activators, or about the *in vivo* consequences of the predominant p63 and p73 isotypes, which lack transactivation domains and appear to have dominant-negative properties. The apparent requirement for p73 in such processes as pheromone detection, neurogenesis and fluid dynamics underscores the role of p73 in mechanisms of sensing environmental and homeostatic stimuli that may in turn represent paradigms adopted by p53 for sensing DNA damage and other cellular stresses. □

Methods

Targeted disruption of p73 gene in embryonic stem cells

A 16.5 kb phage clone containing the 3' end of intron 4 to exon 10 was isolated from a Lambda DASH-II 129/SvJae genomic library and subcloned into pZero vector (Invitrogen). To generate the targeting vector, a 1.4 kb *NheI*–*NheI* fragment, corresponding to exon 5, a portion of exon 6 and adjacent intronic sequences, was deleted and replaced with the neomycin drug-resistance (*NEO*^R) gene driven by the mouse phosphoglycerol kinase (PGK1) promoter and linked to the PKG1 poly(A) sequences. The targeting vector contained ~5.5 kb and 2.5 kb of homology upstream and downstream, respectively, of the *PGK*–*Neo*^R gene and the *pMC-TK* gene³ at its most 5' end. The vector was linearized using a unique *NorI* site and electroporated into the J1 ES cell line¹⁰. G418 and FIAU-resistant ES cell colonies were picked and expanded. Two distinct ES cell lines heterozygous for the disrupted allele were microinjected into blastocysts from B57BL/6 and BALB/c mice, and gave rise to germline transmission of the p73 mutation.

In situ hybridization

Frozen sections were post-fixed for 45 min in 4% paraformaldehyde at 4 °C, washed twice for 5 min in PBS at 4 °C, washed once for 5 min in PBS at room temperature, and dehydrated in 95% ethanol for 1 min. Oligonucleotides were 3'-end labelled with ³⁵S-dATP (Amersham) by terminal transferase (Promega). Radiolabelled probes were added to hybridization solution (10% dextran sulphate, 50% formamide, 1 × Denhardt's, 4 × SSC, 200 mM dithiothreitol (DTT), 250 µg/ml poly(A)) and incubated overnight at 42 °C. Stringency washes were conducted by shaking the slides at 55 °C with 1 × SSC, 0.05% sarkosyl (SK), and 10 mM DTT. For the specific oligonucleotides, see Supplementary Information.

DNA transfection and luciferase assays

SK-NA-S (human neuroblastoma) cells were transfected with p53–luciferase reporter and p73 expression plasmids using Lipofectamine (Life Technologies). In a typical experiment, 1 µg of reporter plasmid p53RGC-Luc was co-transfected with 70 ng of pcDNA3 containing p73α cDNA and 70, 350 or 700 ng of pcDNA3 containing Δ N-p73α cDNA. The final DNA concentration was adjusted to 1.77 µg per well using pcDNA3. Twenty-four hours after transfection, luciferase activity was measured in cell lysates with the Dual Luciferase Reporter Assay System (Promega).

Gel-shift assays

p73 α , p73 α (R293H) or Δ N-p73 α cDNA containing pcDNA-3 were transcribed and subsequently *in vitro* translated using TNT7-coupled reticulocyte lysate system (Promega). Each DNA-binding reaction contained 2.5 μ l of reticulocyte lysate, 2.5 mM of DDT, 10 ng of ³²P-labelled DNA, 1 μ g of salmon DNA as non-specific competitor, 12 μ l of glycerol and Tris-buffered saline (TBS; 25 mM Tris pH 7.5, 130 mM NaCl, 3 mM KCl) to 10 μ l final volume. Reactions were incubated at 23 °C for 30 min, cooled to 4 °C and electrophoresed in 4% non-denaturing polyacrylamide gel in a low salt buffer (0.4 $\times$ TBE).

Received 14 October 1999; accepted 19 January 2000.

1. Kaghad, M. *et al.* Monoallelically expressed gene related to p53 at 1p36, a region frequently deleted in neuroblastoma and other human cancers. *Cell* **90**, 809–819 (1997).
2. Kinzler, K. W. & Vogelstein, B. Life (and death) in a malignant tumour. *Nature* **379**, 19–20 (1996).
3. Ko, L. J. & Prives, C. p53: puzzle and paradigm. *Genes Dev.* **10**, 1054–1072 (1996).
4. Levine, A. J. p53, the cellular gatekeeper for growth and division. *Cell* **88**, 323–331 (1997).
5. Yang, A. *et al.* p63, a p53 homolog at 3q27–29, encodes multiple products with transactivating, death-inducing, and dominant-negative activities. *Mol. Cell* **2**, 305–316 (1998).
6. Yang, A. *et al.* p63 is essential for regenerative proliferation in limb, craniofacial and epithelial development. *Nature* **398**, 714–718 (1999).
7. Mills, A. A. *et al.* p63 is a p53 homologue required for limb and epidermal morphogenesis. *Nature* **398**, 708–713 (1999).
8. De Laurenzi, V. *et al.* Two new p73 splice variants, gamma and delta, with different transcriptional activity. *J. Exp. Med.* **188**, 1763–1768 (1998).
9. Capecchi, M. R. Altering the genome by homologous recombination. *Science* **244**, 1288–1292 (1989).
10. Li, E., Bestor, T. H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915–926 (1992).
11. Donehower, L. A. Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours. *Nature* **356**, 215–221 (1992).
12. Jacks, T. *et al.* Tumor spectrum analysis in p53-mutant mice. *Curr. Biol.* **4**, 1–7 (1994).
13. Kim, K. C. *et al.* Airway goblet cell mucin: its structure and regulation of secretion. *Eur. Respir. J.* **10**, 2644–2649 (1997).
14. Matsui, H. *et al.* Evidence for periciliary liquid layer depletion, not abnormal ion composition, in the pathogenesis of cystic fibrosis airways disease. *Cell* **95**, 1005–1015 (1998).
15. Cressman, V. L., Hicks, E. M., Funkhouser, W. K., Backlund, D. C. & Koller, B. H. The relationship of chronic mucin secretion to airway disease in normal and CFTR-deficient mice. *Am. J. Respir. Cell. Mol. Biol.* **19**, 853–866 (1998).
16. Go, K. G. The normal and pathological physiology of brain water. *Adv. Tech. Stand. Neurosurg.* **23**, 47–142 (1997).
17. Lindeman, G. J. A specific, nonproliferative role for E2F-5 in choroid plexus function revealed by gene targeting. *Genes Dev.* **12**, 1092–1098 (1998).
18. Stanfield, B. B. & Cowan, W. M. The morphology of the hippocampus and dentate gyrus in normal and Reeler mice. *J. Comp. Neurol.* **185**, 393–422 (1979).
19. Rakic, P. & Caviness, V. S. Jr Cortical development: view from neurological mutants two decades later. *Neuron* **14**, 1101–1104 (1995).
20. Ogawa, M. *et al.* The reeler gene-associated antigen on Cajal–Retzius neurons is a crucial molecule for laminar organization of cortical neurons. *Neuron* **14**, 899–912 (1995).
21. D'Arcangelo, G. *et al.* A protein related to extracellular matrix proteins deleted in the mouse mutant reeler. *Nature* **374**, 719–723 (1995).
22. Frotscher, M. Cajal–Retzius cells, Reeler, and the formation of layers. *Curr. Opin. Neurobiol.* **8**, 570–575 (1998).
23. D'Arcangelo, G. & Curran, T. Reeler: new tales on an old mutant mouse. *BioEssays* **20**, 235–244 (1998).
24. Alcantara, S. *et al.* Regional and cellular patterns of reelin mRNA expression in the forebrain of the developing and adult mouse. *J. Neurosci.* **18**, 7779–7799 (1998).
25. McEwen, B. S. Stress and hippocampal plasticity. *Annu. Rev. Neurosci.* **22**, 105–122 (1999).
26. Gage, F. H., Kempermann, G., Palmer, T. D., Peterson, D. A., Ray, J. Multipotent progenitor cells in the adult dentate gyrus. *J. Neurobiol.* **36**, 249–266 (1998).
27. Bargmann, C. I. Olfactory receptors, vomeronasal receptors, and the organization of olfactory information. *Cell* **90**, 585–587 (1997).
28. Tirindelli, R. & Mugnat-Caretta, C., Ryba, N. J. Molecular aspects of pheromonal communication via the vomeronasal organ of mammals. *Trends Neurosci.* **21**, 482–486 (1998).
29. Matsunami, H. & Buck, L. A multigene family encoding a diverse array of putative pheromone receptors in mammals. *Cell* **90**, 775–784 (1997).
30. Herrada, G. & Dulac, C. A novel family of putative pheromone receptors in mammals with a topographically organized and sexually dimorphic distribution. *Cell* **90**, 763–773 (1997).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We would like to thank F. Borriello, H. Green, C. Westphal, P. Ferrara, T. Rapoport and L. Buck for helpful discussions; L. Du for blastocyst injections; A. Goffinet for the reelin antibody; H. Liu for the mouse genomic library; and J. Williams for histology preparation. This work was supported by the American Cancer Society and the Council for Tobacco Research (F.M.), and the NIH (R.B., A.S., P.D. and F.M.).

Correspondence and requests for materials should be addressed to F.M. (e-mail: fmckeeon@hms.harvard.edu) or D.C. (e-mail: daniel.caput@wanadoo.fr). GenBank accession codes for *M. musculus* p73 α and *M. musculus* Δ N-p73 α mRNAs are Y19234 and Y19235, respectively.

Single-molecule studies of the effect of template tension on T7 DNA polymerase activity

Gijs J.L. Wuite*, Steven B. Smith*, Mark Young†, David Keller‡ & Carlos Bustamante*

* Department of Physics and Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA

† Institute of Molecular Biology, University of Oregon, Eugene, Oregon 97403, USA

‡ Department of Chemistry, University of New Mexico, Albuquerque, New Mexico 87131, USA

T7 DNA polymerase^{1,2} catalyses DNA replication *in vitro* at rates of more than 100 bases per second and has a 3'→5' exonuclease (nucleotide removing) activity at a separate active site. This enzyme possesses a 'right hand' shape which is common to most polymerases with fingers, palm and thumb domains^{3,4}. The rate-limiting step for replication is thought to involve a conformational change between an 'open fingers' state in which the active site samples nucleotides, and a 'closed' state in which nucleotide incorporation occurs^{3,5}. DNA polymerase must function as a molecular motor converting chemical energy into mechanical force as it moves over the template. Here we show, using a single-molecule assay based on the differential elasticity of single-stranded and double-stranded DNA, that mechanical force is generated during the rate-limiting step and that the motor can work against a maximum template tension of

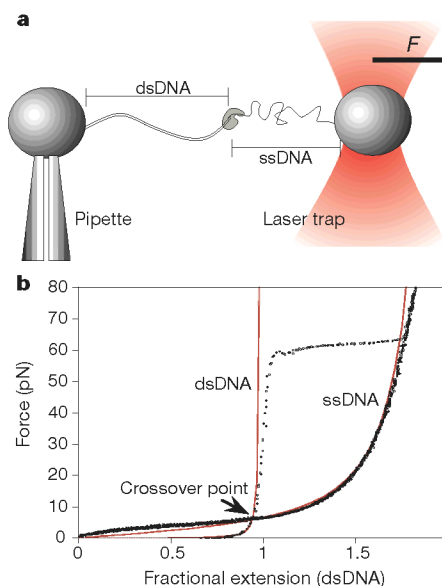


Figure 1 Optical trap setup. **a**, A 10,416-base pair plasmid DNA fragment was prepared as described¹⁶ and attached between two beads, one held on the tip of a glass pipette, the other in an optical trap. Single-stranded DNA was obtained by using the force-induced exonuclease activity of T7 DNA polymerase to remove any desired length of the non-template strand. The end-to-end length of the DNA was obtained by video imaging of the bead positions, and the force (F) was measured using the change in light momentum which exits the dual-beam trap⁶. **b**, Force–extension data for dsDNA and ssDNA (dotted lines), compared with the wormlike chain model using ssDNA and dsDNA persistence lengths of 0.7 nm and 53 nm respectively (solid lines)^{7,8}. Difference between ssDNA and WLC curves at low tension is due to partial hairpin formation and disappears if magnesium is removed from buffer.

~34 pN. Estimates of the mechanical and entropic work done by the enzyme show that T7 DNA polymerase organizes two template bases in the polymerization site during each catalytic cycle. We also find a force-induced 100-fold increase in exonucleolysis above 40 pN.

We measured T7 DNA polymerase (DNAP) activity by using the optical-trap shown in Fig. 1a⁶. Because single-stranded (ss) and double-stranded (ds) DNA differ in length at any given tension^{6–8} (Fig. 1b), conversion between these two forms changes the molecule's tension if the end-to-end distance of the template is held constant. Alternatively, the molecule's end-to-end distance changes if the tension is held constant. The end-to-end distance of ssDNA (Fig. 1b) is shorter than that of dsDNA for tensions below 6.5 pN ('crossover point') because ssDNA, despite having about twice the contour length of dsDNA, is more retractile owing to its greater flexibility. Above 6.5 pN, however, contour length predominates over entropy and ssDNA is longer than dsDNA. The force–extension curve of a molecule that is partly ssDNA and partly dsDNA can be fit to a linear combination of ssDNA and dsDNA stretching curves⁶. Thus, the progress of DNAP can be followed by the number of single-stranded bases, N_{ss} , remaining in the template at time t

$$N_{ss}(t) = \frac{x_{meas}(F, t) - x_{ds}(F)}{x_{ss}(F) - x_{ds}(F)} * N_{tot} \quad (1)$$

where x_{meas} is the end-to-end distance of the molecule at force F ; $x_{ds,ss}(F)$ are the end-to-end distances of fully double- or single-stranded DNA at that force, and N_{tot} is the total number of bases in the template.

Figure 2a (upper line) plots the ssDNA fraction remaining at time t as DNAP replicates against an applied tension of 20 pN. The instantaneous polymerization rate (lower line) determined from the time derivative of this curve shows bursts of activity. Two lines of evidence suggest that each burst corresponds to a DNAP molecule loading onto the 3' end of the growing chain, replicating proces-

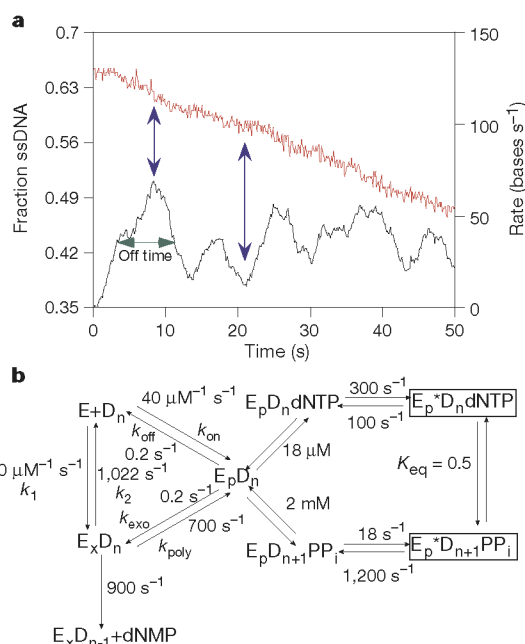


Figure 2 Polymerization kinetics. **a**, Replication of a ssDNA template under 20 pN tension using 8 nM DNA. Upper curve, conversion to dsDNA plotted as fraction of ssDNA left in the template versus time. Lower curve, polymerization rate obtained by differentiating upper curve after smoothing it with a moving-average filter (24 data points) to reduce Brownian noise. **b**, Induced fit kinetic pathway of T7 DNAP according to Patel *et al.*⁹ with k_1 and k_2 taken from Donlin *et al.*¹⁶. States of the DNA–enzyme complex are shown as: E, enzyme in solution; E_p, DNA bound in polymerase active site or exo active site; D_n, DNA primer with length n ; dNTP, deoxynucleotide; and PP_i, pyrophosphate.

sively, and falling off. First, the mean width of a burst is force and concentration independent, and the rate corresponding to this width ($0.13 \pm 0.1 \text{ s}^{-1}$; $N = 62$) is near the off rate from the polymerization site measured in bulk⁹. Second, varying the DNAP concentration changed the width of the gaps between bursts. At 0.8 nM, gap widths average $7 \pm 4 \text{ s}$, consistent with $\sim 7 \text{ s}$ calculated from the enzyme loading rates from solution ($\sim 180 \text{ s}^{-1} \mu\text{M}^{-1}$; Fig. 2b). Increasing the DNAP concentration to 8 nM, decreased gap size to $1.7 \pm 0.8 \text{ s}$. At 80 nM and 880 nM, however, this trend reversed and the gaps increased to $7 \pm 2 \text{ s}$ and $50 \pm 26 \text{ s}$, respectively. Notably, T7 DNAP can bind non-specifically along ssDNA with a dissociation constant of $\sim 800 \text{ nM}$ ¹⁰. Therefore, such binding may block enzyme reloading at the 3' end, increasing gap widths. Because the average burst height remained independent of enzyme concentration (data not shown), non-specific binding does not seem to impede translocation once an enzyme is specifically bound. Occasionally, replication would stop for extended periods (up to 30 min), indicating 'roadblocks' that the polymerase would not cross. Causes might include exogenous DNA hybridized to the template or bases missing from the template because of chemical or enzymatic damage. Template hairpins are an unlikely cause as blockage persists above 15 pN, where such structures should pull out¹¹.

Figure 2a shows surprising diversity in polymerization rates among individual DNAP molecules. Burst heights typically vary between 26 and 60 bases s^{-1} ($N = 39$, s.d. = 17), at a template tension of 20 pN. As the intrinsic rate of each DNAP molecule can be determined from analysis of burst heights (see Methods), such variations probably reflect differences in enzymatic activity among individual molecules. Similar differences have been reported for other enzymes^{12,13}.

The effect of template tension on the polymerization rate was determined either by holding the tension constant through force-feedback or by holding the template strand at constant end-to-end length. This length was chosen beyond the crossover point, so that polymerization increased template tension until the system halted itself, having sampled all intermediate tensions (Fig. 3). At the lowest tension, the replication rate of T7 DNAP was $\sim 100 \text{ bases s}^{-1}$. Raising template tension increased the replication rate until a maximum of $\sim 200 \text{ bases s}^{-1}$ was reached at about 6 pN. Further increase in tension, however, caused the rate to decrease until polymerization stalled. For 12 constant-distance runs, the mean stall force was $34 \pm 8 \text{ pN}$. On rare occasions, polymerization

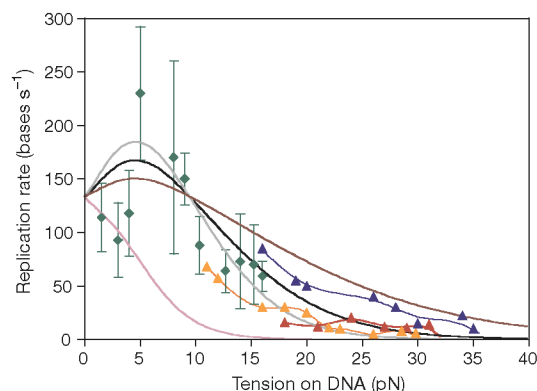


Figure 3 Polymerization rate versus template tension. Diamonds represent 50 polymerization bursts taken at 11 different tensions (error bar, s.d.); triangles represent three traces fitted through a succession of replication bursts measured at constant end-to-end distances until a stalling force is reached; thick lines represent fits of equation (4) using $a = 0$, $k_0 = 130 \text{ bases s}^{-1}$, and $n = 1$ (brown), $n = 2$ (black), $n = 3$ (grey) and $a = 1 \text{ nm}$, $n = 2$ (purple). Noise increases near 0 pN and the crossover point, because the numerator and denominator of equation (1) become small while the brownian motion does not decrease.

proceeded briefly above 50 pN. Replication did not restart spontaneously at these high tensions even after several minutes, but replication always resumed after lowering the tension. The force would then rise once again until a new stalling force, usually different from the previous one, was reached, perhaps reflecting the effect of template sequence, or the stochastic nature of the stalling process itself.

The sensitivity of the polymerization rate even to low tensions indicates that the rate-limiting step is directly affected by force. This force dependence is consistent with the induced-fit model of Wong *et al.*¹⁴, in which the enzyme changes conformation during the rate-limiting step. Crystal structures of the closed state show the fingers rotated $\sim 40^\circ$ to align the different components in the active site^{3,4}, and the ssDNA bending sharply in relation to the primer as it leaves this site.

If the tension, F , on the template can exert a torque on the fingers, the work done by the enzyme would include a term aF , where a is the distance moved by the fingers along the pulling direction of the template. Further, assume that closing the fingers organizes n adjacent sugar-phosphate units from single- to double-stranded geometry, $n - 1$ of which are released when the fingers reopen. Then the work W to close the fingers is

$$W(F) = aF + nF(x_{ss}(F) - x_{ds}(F)) \quad (2)$$

where $x_{ss}(F)$ and $x_{ds}(F)$ are the end-to-end distances per base of ssDNA and dsDNA at tension F . The second term in equation (2) changes sign at 6.5 pN, aiding or opposing replication below or above this force and causing the instrument to do work on the reaction or the reaction to do work on the instrument, respectively. Closing of the fingers also clamps the template strand reducing its degrees of freedom. Tensions applied to the template decrease its entropy and, consequently, the energetic cost of closing the fingers, thus speeding up the reaction. The entropy change, ΔS , to convert ssDNA into dsDNA at any given force can be obtained from the areas under the experimental force-extension curves (Fig. 1b), that is,

$$T\Delta S(F) = n \left[\int_0^{x_{ss}(F)} F_{ss} dx - \int_0^{x_{ds}(F)} F_{ds} dx \right] \quad (3)$$

where $F_{ds,ss}$ are the experimental forces required to extend the chains by an amount x . If we assume that the terms in equations (2) and (3) contribute to the activation energy required to reach a transition state (for example, the complex with fingers half-closed) from the open state, then the rate coefficient (k) for this step is

$$k = k_0 e^{-((w - T\Delta S)/k_B T)} \quad (4)$$

where k_0 is the rate coefficient at zero force; k_B is the Boltzmann constant and T is the temperature. Figure 3 compares equation (4) to the data using various values of a and n . The best fit is obtained for $a = 0$ and $n = 2$, indicating that template tension exerts little torque opposing finger closure and that two adjacent sugar-phosphate units are organized by the fingers in this process. These results are supported by the structure of the closed complex which shows the template strand avoiding the finger tips and passing around the side of the fingers through a shallow cleft^{3,4}. Moreover, two adjacent template bases appear immobilized in this structure. The first, opposite the incoming nucleotide, adopts a B-form structure, whereas the second is kinked outward, almost perpendicular to the pulling direction. The interphosphate distance corresponding to these two bases is close to that of dsDNA. Subsequent bases ($n > 2$) appear disorganized in the structure. Single-molecule studies of an exonuclease-deficient mutant of T7 DNAP (Sequenase) also suggest immobilization of two bases during finger closure (B. Maier, D. Bensimon and V. Croquette, personal communication).

When template tension was increased above 40 ± 3 pN ($N = 16$), a fast exonucleolysis (30 ± 11 bases s^{-1}) was initiated with or

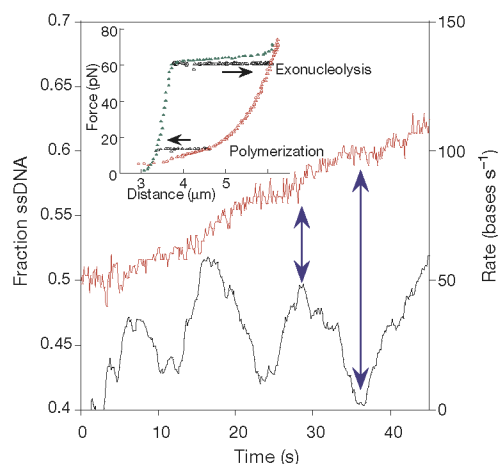


Figure 4 Exonuclease digestion of primer strand by T7 DNAP (at 8 nM) at template tension of 50 pN. Upper curve, ssDNA fraction of 10-kilobase template. Lower curve, exonuclease rate obtained by differentiating upper curve after application of a 3-s moving-average filter. Inset, force-extension curve for a dsDNA molecule (green points) before it was almost entirely converted to ssDNA by exonucleolysis at a constant force of 60 pN (upper horizontal points, right arrow). At the end of this process a force-extension curve for the ssDNA was obtained (red points). Finally, in the presence of dNTPs, the tension was decreased to 15 pN to allow T7 DNAP to reconvert ssDNA into dsDNA (lower horizontal points, left arrow).

without dNTPs present (Fig. 4). Decreasing tension below ~ 34 pN caused exonucleolysis to halt and polymerization to resume. Switching between these opposite activities (inset, Fig. 4) could be repeated many times on one template (data not shown). Exonucleolysis force dependence was measured either using constant force or constant end-to-end distance (Fig. 5). The exonucleolysis rate became force independent above 42 pN and is ~ 100 times faster than observed at zero tension on dsDNA, where the exo rate is limited by the escape of the 3' end from the polymerization site (Fig. 2b). Force-induced exonucleolysis appears as bursts of activity (Fig. 4, lower line). Presumably, template tension shifts the equilibrium in favour of exonucleolysis either by increasing the escape rate from (k_{off} and k_{exo}) or decreasing the binding rate to (k_{on} and k_{poly}) the polymerization site, or both. However, if the escape rate from the polymerization site were increased by two orders of magnitude, such escape would still be rate limiting for exonucleolysis ($0.2 s^{-1} \times 100$) which would appear continuous at our temporal resolution. The gaps observed in the exonuclease rate are instead consistent with a 100-fold decrease in binding rate to the polymerization site. Because the enzyme bound through its exonuclease site associates/dissociates rapidly from solution to the 3' end (Fig. 2b), many DNAPs can bind, exonucleate and dissociate, before one of them moves back to the polymerase site. It then takes several seconds for this enzyme to escape the polymerase site ($k_{exo} = 0.2 s^{-1}$), resulting in gaps of activity as observed in Fig. 4. Thus, each exo-pause results from one DNAP molecule lingering in the polymerization active site, while each exo-burst, seeming continuous because of our temporal resolution, could result from the action of many individual DNAP molecules.

Although the gaps can be explained by a slow escape of the 3' end from the poly site, the heights of the exo-bursts are inconsistent with published association/dissociation rates of the exo site from solution (k_1 and k_2 , Fig. 2b). Unexpectedly, the forced-induced exonucleolysis rate remains relatively constant for DNAP concentration between 800 nM and 8 nM, dropping by 50% only when the concentration is lowered to 0.8 nM. Under zero tension, it is thought that several base pairs must be melted to allow binding of the DNA primer strand to the exonuclease site⁵, and forces greater

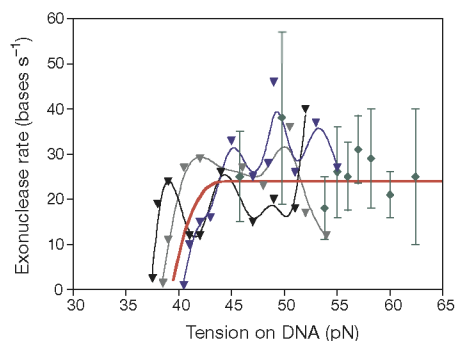


Figure 5 Force dependence of 3'→5' exonuclease reaction. Diamonds represent average rates for 49 (total) exonuclease bursts measured at 9 different forces. Traces represent three lines fitted through successions of exonucleolysis burst heights (triangles), initiated at high tensions, on DNAs kept at constant end-to-end distances. Digestion lowers the tension in the DNA until the fast exonucleolysis stops. The upper limit for these experiments is determined by the overstretching force, 65 pN^{6,19}. Near the stalling force for polymerization, a competition was occasionally observed between exonucleolysis and polymerization which caused the template tension to bounce up and down every few seconds.

than 40 pN are known to promote fraying in dsDNA¹⁵. Perhaps the exonucleolysis rates simply reflect the rate of fraying of DNA at the tensions applied to the template and the known fast exo activity of T7 DNAP¹⁶ on ssDNA. To test this idea, we investigated the effect of template tension on the activity of *Escherichia coli* exonuclease I. This enzyme attacks only ssDNA¹⁷ at its 3' end, so it should not digest the primer strand unless it frays. No activity was detected under a tension of 40 pN, but at 50 pN bases were removed at a rate of $\sim 200 \text{ s}^{-1}$. This fast exonucleolysis suggests that 3' end fraying is not rate limiting during exonucleolysis by T7 DNAP. Thus, the force-induced exonucleolysis initiated at 40 pN is probably a specific property of T7 DNAP itself. For example, 40-pN tension may deform the dsDNA geometry at the 3' end enough to trigger the enzyme's proof-reading function. But, as Fig. 5 shows, the fraying is rate limited in the presence of T7 DNAP. Perhaps the melting rate is mediated by some interaction with the enzyme which is independent of the tension. Future single-molecule experiments should elucidate the mechanism of force-induced exonucleolysis and provide additional insight into the mechanochemistry of DNA polymerase. □

Methods

Single-molecule assay

T7 DNAP (T7 gene 5 in a 1:1 complex with thioredoxin) was used in various concentrations (0.8–880 nM), where 1 nmol = ~ 230 activity units (Amersham). Replication buffer was 40 mM Tris pH 7.5, 5 mM MgCl₂, 50 mM NaCl, 50 $\mu\text{g ml}^{-1}$ BSA, 0.1% NaN₃, 5 mM dithiothreitol and 0.6 mM (each) dNTPs. The end of one of the DNA chains was covalently coupled to a bead while the other end of the same chain was attached

through a biotin/streptavidin linkage to the second bead¹⁸ (Fig. 1a). Exonuclease I (USB) was used at 50 units ml⁻¹ in replication buffer altered to pH 8 and without dNTPs.

Temporal resolution

Data were collected at 8 Hz with a bandwidth of 60 Hz. The polymerization rate data was averaged over 3 s and, therefore, the height of bursts longer than 3 s should represent the intrinsic activity of individual molecules. For the expected average replication time (~ 5 s), most burst heights will be accurately determined with this temporal resolution. No correlation between burst height and width (measured at the half maximum) was found for bursts longer than 3 s.

Processivity

Because the replication rate varies with force (Fig. 3), whereas the dissociation rate (k_{off}) seems to remain constant, the processivity varies with force and is ~ 420 bases (60 bases s⁻¹, 0.13 s⁻¹) at 15 pN.

Received 3 September 1999; accepted 6 January 2000.

- Modrich, P. & Richardson, C. C. Bacteriophage T7 deoxyribonucleic acid replication *in-vitro*. Bacteriophage T7 DNA polymerase: an enzyme composed of phage- and host-specific subunits. *J. Biol. Chem.* **250**, 5515–5522. (1975).
- Tabor, S., Huber, H. E. & Richardson, C. C. *Escherichia coli* thioredoxin confers processivity on the DNA polymerase activity of the gene 5 protein of bacteriophage T7. *J. Biol. Chem.* **262**, 16212–16223 (1987).
- Doublie, S. & Ellenberger, T. The mechanism of action of T7 DNA polymerase. *Curr. Opin. Struct. Biol.* **8**, 704–712 (1998).
- Doublie, S. *et al.* Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. *Nature* **391**, 251–258 (1998).
- Johnson, K. A. conformational coupling in DNA polymerase fidelity. *Annu. Rev. Biochem.* **62**, 685–713 (1993).
- Smith, S. B., Cui, Y. & Bustamante, C. Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. *Science* **271**, 795–799 (1996).
- Marko, J. F. & Siggia, E. D. Stretching DNA. *Macromolecules* **28**, 8759–8770 (1995).
- Bustamante, C., Marko, J. F., Siggia, E. D. & Smith, S. B. Entropic elasticity of lambda-phage DNA. *Science* **265**, 1599–1600 (1994).
- Patel, S. S., Wong, I. & Johnson, K. A. Pre-steady-state kinetic analysis of processive DNA replication including complete characterization of an exonuclease-deficient mutant. *Biochemistry* **30**, 511–525 (1991).
- Huber, H. E., Tabor, S. & Richardson, C. *Escherichia coli* thioredoxin stabilizes complexes of bacteriophage T7 DNA polymerase and primed templates. *J. Biol. Chem.* **262**, 16224–16232 (1987).
- Essevaz-Roulet, B., Bockelmann, U. & Heslot, F. Mechanical separation of the complementary strands of DNA. *Proc. Natl Acad. Sci. USA* **94**, 11935–11940 (1997).
- Xue, Q. & Yeung, E. Differences in the chemical reactivity of individual molecules of an enzyme. *Nature* **373**, 681–683 (1995).
- Lu, H. P., Xun, L. & Xie, X. S. Single-molecule enzymatic dynamics. *Science* **282**, 1877–1882 (1998).
- Wong, I., Patel, S. S. & Johnson, K. A. An induced-fit kinetic mechanism for DNA replication fidelity: direct measurement by single-turnover kinetics. *Biochemistry* **30**, 526–537 (1991).
- Gurrieri, S., Smith, S. B. & Bustamante, C. Trapping of megabase-sized DNA molecules during agarose gel electrophoresis. *Proc. Natl Acad. Sci. USA* **96**, 453–458 (1999).
- Donlin, M. J., Patel, S. S. & Johnson, K. A. Kinetic partitioning between the exonuclease and polymerase sites in DNA error correction. *Biochemistry* **30**, 538–546 (1991).
- Lehman, I. R. & Nussbaum, A. L. On the specificity of *E. coli* exonuclease I. *J. Biol. Chem.* **239**, 2628–2636 (1964).
- Hegner, M., Smith, S. B. & Bustamante, C. Polymerization and mechanical properties of single RecA-DNA filaments. *Proc. Natl Acad. Sci. USA* **96**, 10109–10114 (1999).
- Ciuzel, P. *et al.* DNA: an extensible molecule. *Science* **271**, 792–794 (1996).

Acknowledgements

We thank J. Davenport and M. Hegner for their help with the template and suggestions, and D. Bensimon, B. Maier and V. Croquette for sharing their results before publication.

Correspondence and requests for materials should be addressed to C.B. (e-mail: carlos@alice.berkeley.edu).

Equipped for the purpose

A trawl through general laboratory equipment.

Scotsman AF 80

Hubbard Refrigeration Ltd

www.hubbardref.co.uk

Keep your flaker by your side

The Scotsman AF 80 is a compact (823 mm high × 529 mm wide × 525 mm deep) flaked-ice machine with an integrated storage bin, designed for small-scale use in research laboratories. This 'personal flaker' requires no ventilation, so it can be sited conveniently alongside other units, providing a steady supply of ice without the need to rely on a large central machine. The AF 80 can produce 70 kg of ice per day, and holds up to 27 kg of ice in its integral storage bin.

Reader Service No. 100

Chemware utensils

Norton Performance Plastics

www.nortonplastics.com

High chemical resistance in lab plastics

Chemware laboratory utensils have a PTFE fluoropolymer coating that is chemically inert, with a low coefficient of friction, making them especially suited to work requiring precise control, ultra-purity and chemical resistance. Spatulas are made of either stainless steel or a nickel-stainless-steel alloy, and are available in Hayman, double flat-end, lab spoon and palette-knife styles. Forceps are available for handling materials that must contact corrosive chemicals and are also available in an all-PTFE style (with an encapsulated metal insert) for use at temperatures from -450 °F to +550 °F. Tongs will not scratch glassware and are made for larger handling jobs. Norton recently combined operations with Furon Company under the name Saint-Gobain Performance Plastics.

Reader Service No. 101

Miniature spectrometer

Anglia Instruments www.angliainst.co.uk

An alternative to conventional Raman analysers

This miniature Raman spectroscopy system, a joint venture between Ocean Optics and Boston Advanced Technologies, has been launched in the UK by Anglia Instruments. It is especially suited to reaction monitoring, product identification, remote sensing and the characterization of highly scattering particulate matter in aqueous solutions, gels and other media. Utilizing a high-output 500-mW solid-state laser to excite the sample, the R-2000 collects the backscattering generated with an SMA-terminated fibre-optic probe, the signal then being detected and analysed by a sensitive

2048 CCD array spectrometer. An ISA-bus or A/D converter provides the interface with a desktop or notebook PC, which runs R-2000 Windows software for real-time acquisition and display of spectral data.

Reader Service No. 102

CD spectropolarimeters

Jasco (UK) Limited

www.jasco.co.uk

Extended wavelength range

The J-800 series of circular dichroism spectropolarimeters incorporate new hardware and software functions in a compact design. Four-channel data acquisition is standard and measurement modes include CD, FDCD, ORDE, ORDM, LD, UV and fluorescence. There are three scanning modes for continuous, step and autoreponse analysis. The wavelength range is extended from 163 to 900 nm (with an option to run to 1,100 nm). A fluorescence attachment, automatic titrator and stopped-flow device are also available.

Reader Service No. 103

Airflow controller

Trox Technik

www.trox.de

Something in the air

The Labcontrol airflow controller monitors and controls the airflow in laboratory rooms and fume cupboards inexpensively by supplying only enough air to fume cupboards to maintain sash velocity — a lower

sash position reduces the air flow rates. The Labcontrol incorporates a monitoring device which precisely calculates the air volume and maintains the sash velocity required. Trox Labcontrol is designed to be safe in operation with thermal loads such as Bunsen burners. Other features include a plastic attenuator which is non-corrosive and contributes to low noise levels.

Reader Service No. 104

Lab measurement

Semat

www.semat.co.uk

Multiple-parameter laboratory measurements

WTW inoLab meters, available from Semat, are designed for daily routine measurements of pH, oxygen, conductivity, salinity and TDS, as well as temperature. They come in three versions: Level 1, which includes a routine bench-top laboratory meter with battery or a.c. power which can be used as a bench-top instrument or wall-mounted. Level 2 includes a precision meter, with built-in measurement storage for GLP-conform documentation, a bi-directional RS232 digital interface and optional built-in printer. Level 3 is a high-precision meter with graphics display, digital recorder and PC software.

Reader Service No. 105

UV Mini

Shimadzu Scientific Instruments

www.shimadzu.com

Compact UV-Vis spectrophotometer

The new UV-Vis spectrophotometer, the UV *Mini*, operates in photometric, spectrum and quantitation modes. Optional program packs are available to expand instrument functionalities, including quantification of DNA, proteins and enzyme kinetics. A universal sample compartment accommodates accessories, which are compatible with all Shimadzu UV-Vis instruments. Data Manager software provides unlimited data storage with expanded archiving and management capabilities.

Reader Service No. 106

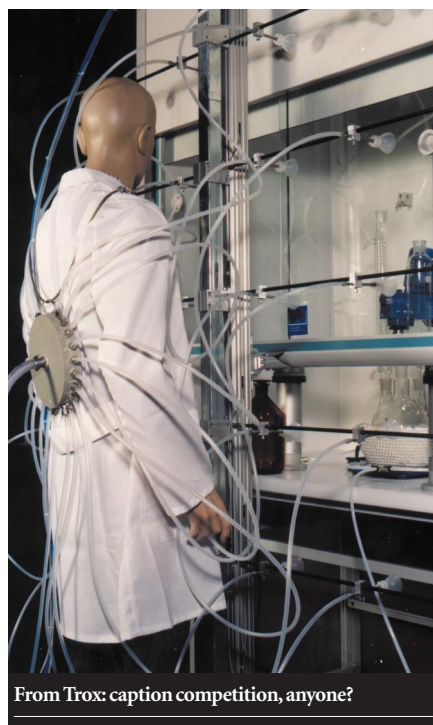
Genesys 5

Spectronic Unicam EMEA

www.spectronic.com

Precise positioning of sample cells

The Spectronic Genesys 5 UV-Vis spectrophotometer has a 5-nm bandwidth over the 200–1,100-nm wavelength range and contains an 8-position cell holder for precise, automatic positioning of sample cells. The



From Trox: caption competition, anyone?

new on the market

built-in graphic LCD screen provides the operator with details of the test being performed, including application program type, analytical wavelength, sample ID, test name and test results. Results may be sent to the optional internal printer or to external devices via the built-in RS232C- or Centronics-compatible ports. Methods and data can be stored on industry-standard PCMCIA softcards that are transferable to other Genesys 5 instruments and eliminate the need for manual record-keeping.

Reader Service No. 107

Fluorescence spectrophotometer

Varian Instruments www.varianinc.com

Flexible life-science measurements

A new mid-range fluorescence instrument, the Cary Eclipse fluorescence spectrophotometer, was introduced last November. The Eclipse is suitable for a range of life-science measurements in biochemical and pharmaceutical laboratories, in applications such as purifying solvents for combinatorial chemistry. It can also be used to identify petrochemical mixtures using graphical representations of data, including contour and three-dimensional plots. Windows-based software includes the spectroscopy Applications Development Language (ADL). Collection mode can be set to fluorescence, phosphorescence or chem/bioluminescence.

Reader Service No. 108

Handheld homogenizer

Misonix www.misonex.com

Homogenization at your fingertips

The Microson Ultrasonic Homogenizer is a high-power, handheld, thumb-activated homogenizer for general research use. Applications include cell disruption, particulate dispersion, mixing, homogenizing and focused cleaning of intricate instruments. The unit features automatic tuning, variable



Mixed blessing: the Microson at work.

amplitude control, and a digitally displayed wattmeter. The Microson measures $7.5 \times 13 \times 27$ inches and comes with a 5-inch-long titanium alloy probe that can take a tip diameter of between $\frac{3}{32}$ and $\frac{1}{4}$ inches.

Reader Service No. 109

FieldSpec Pro

Analytical Spectral Devices www.asdi.com

Spectroradiometry on the move

ASD suggest applications varying from remote sensing and precision agriculture to oceanography and environmental assessment for the FieldSpec Pro. This portable UV/Vis/NIR spectroradiometer measures $13 \times 4.5 \times 16$ inches and weighs 16 pounds, so it's small enough to be mounted on a backpack. The incorporated carrier detaches from the shoulder straps to fit easily into



FieldSpec Pro: take your work with you.

the overhead compartment of an aircraft and it meets all international airline carry-on regulations.

Reader Service No. 110

Fiberlight

Heraeus Noblelight Ltd

www.heraeus-noblelight.com

Handheld deuterium illumination

Developed for use with handheld, fibre-coupled UV-Vis spectrometers, Fiberlight features a miniature deuterium lamp for UV light and a tungsten lamp for visible light, with an integral shine-through arrangement and optical system. UV light from deuterium lamps is used for optical analysis in spectroscopy and high-pressure liquid chromatography, and deuterium lamps are used for these applications because of their high stability and line-free, continuous spectrum in the UV region. Fiberlight allows a deuterium lamp to be used in handheld analysis.

Reader Service No. 111

Temperature data loggers

Anville Instruments

www.anvilleinstruments.co.uk

Integrated, expandable

The Series 425/TC has direct type-T thermocouple inputs via compensated connectors, eliminating cold-junction errors and enabling these devices to measure type-T inputs over a range from -100°C to $+200^\circ\text{C}$ with a system accuracy of $\pm 0.25^\circ\text{C}$ and resolution of 0.01°C . It comes with Scan 1000 X425 software. Its 16 analogue inputs can be expanded, daisy-chain fashion, by up to 8 further units using X425 software. A DDE link to Microsoft Excel allows online F_0 calculations.

Reader Service No. 112

These notes are compiled in the Nature office from information provided by the manufacturers. For more details, fill in the reader service card.

ADVERTISEMENTS

Check out our homepage at <http://www.resgen.com>

**BACs
YACs
PACs
cDNAs**

Research Genetics provides clones and colony membranes for a number of human, mouse, and rat libraries as well as a custom screening service for the libraries.

Research Genetics, Inc.
U.S. or Canada 800-533-4363
U.K. 0-800-89-1393
FAX 256-536-9016

Look no further
for the best value in
Oligo Synthesis

Competitive pricing
Routine 24 hour dispatch • Every Oligo QC verified by PAGE
Rigorous batch QC by PCR and Mass Spec
Web ordering and tracking • Local Freephone numbers
A great deal in any language!!

SIGMA GENOSYS (+44) (0) 1223 839000 • Fax (+44) (0) 1223 839200
email: info@sigma-genosys.co.uk • www.sigma-genosys.co.uk

READER ENQUIRY NO. 4

READER ENQUIRY NO. 33